REPORT RESUMES

ED 002 849 24
DIFFERENCES BETWEEN GOOD AND POOR PROBLEM SOLVERS.
BY- TATE, MERLE W. AND OTHERS
PENNSYLVANIA UNIV., PHILADELPHIA,SCH. OF EDUCATION
REPORT NUMBER CRP-368 PUB DATE SEP 59
EDRS PRICE MF-$0.18 HC-$4.12 103P.

DESCRIPTORS- *PROBLEM SOLVING, *COMPARATIVE ANALYSIS, *COMPARATIVE TESTING, COMPARATIVE STATISTICS, *GRADE 7, *GRADE 8, CALIFORNIA TEST OF MENTAL MATURITY, DAVIS EELLS GAMES, PHILADELPHIA, PENNSYLVANIA

SEVENTH- AND EIGHTH-GRADE PUPILS WERE SELECTED FOR THIS STUDY OF DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS AND FOR AN ANALYSIS OF PROBLEM-SOLVING ABILITY. THESE PUPILS WERE OLD ENOUGH FOR MANY TESTS IN LARGE-SCALE FACTOR ANALYSES OF REASONING ABILITY, AND HAVE HAD LITTLE PARTICIPATION IN FORMAL PROBLEM-SOLVING ACTIVITIES. SEVERAL TESTS OF REASONING AND PROBLEM-SOLVING ABILITY WERE USED--(1) DIFFERENTIAL APTITUDE TESTS OF VERBAL AND ABSTRACT REASONING, (2) DAVIS-EELLS GAMES, AND (3) THOUGHT PROBLEMS. ALSO, TESTS OF MECHANICAL REASONING AND MATRICES WERE USED AS A VALIDATING RATHER THAN A CRITERION MEASURE. ON THE BASIS OF THESE RESULTS AND PUPIL ACHIEVEMENT RECORDS, 117 EACH OF POOR AND GOOD PROBLEM-SOLVERS WERE SELECTED AS SUBJECTS. THE SECOND BATTERY INCLUDED GROUP TESTS OF MEMORY RECOGNITION OF PROBLEMS, WORD ASSOCIATION, IDEATIONAL FLUENCY, CLOSURE, JUDGMENT, AND PREFERENCES. FIFTY SUBJECTS IN EACH GROUP WERE SELECTED FOR INDIVIDUAL TESTS. GOOD PROBLEM-SOLVERS PERFORMED SIGNIFICANTLY BETTER THAN POOR ONES ON NEARLY ALL TESTS IN WHICH QUALITY OF RESPONSE, ACCURACY, OR JUDGMENT WAS REQUIRED. WITHOUT EXCEPTION, THE SUPERIORITY OF THE GOOD PROBLEM-SOLVERS INCREASED WITH THE COMPLEXITY OF THE TASKS OR WITH THE RESTRICTION OF THE REQUIREMENTS. THE COMPLEX ABILITY OF PROBLEM RECOGNITION AND JUDGMENT SEEM TO BE INDEPENDENT OF INTELLIGENCE OR READING SKILL. IDEATIONAL FLUENCY APPEARED TO FAVOR THE POOR PROBLEM-SOLVERS WHEN TESTS WERE SCORED ON QUANTITY RESPONSE. (GC)

ED 002849

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS

By

Merle W. Tate

Barbara Stanier

Berj Harootunian

School of Education
University of Pennsylvania
Philadelphia, Pennsylvania

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

#368

September 1959

8414A
(219)

PREFACE

This study was carried out under the Cooperative Research Program of the United States Office of Education. We wish to acknowledge first of all the advice and help of two members of that Office: Dr. William Asher and Miss Alice Y. Scates.

Many other persons have assisted with the study. Dr. Dorothy C. Adkins and Dr. Thelma G. Thurstone of the University of North Carolina, Dr. W. A. Botzum of the University of Portland, and Dr. J. P. Guilford of the University of Southern California made available various tests and materials which had been used in factor analyses of reasoning ability.

Mr. Richard Cross, Mr. William Phillips, Mrs. Lillian Rau, and Mr. William Storch of the Abington Junior High School; and Mrs. Elizabeth Barcus, Mr. Douglas Duncan, Mr. Robert Emberger, Mr. Charles Forbes, Mr. Louis Gershman, Mr. Robert Gwinner, Mrs. Helen Hort, Mrs. Gayle How, Miss Mary Elizabeth Hunsicker, Mr. John Rank, Mrs. Eleanor Roberts, Mr. Raymond Spafford, Miss Grace Staley, and Miss Ann Wertz of the Lansdowne-Aldan Junior-Senior High School administered the group tests.

Principal Charles Oister and his assistant, Mrs. June Mussman, of the Abington Junior High School and Principal Joseph Moore and his assistant, Mr. Stanley Kleckner, of the Lansdowne-Aldan Junior-Senior High School worked out the details of the group testing program. Their sustained interest and cooperation contributed greatly to the study.

Mrs. Marcella Bell, Mr. Barry Slepian, Mrs. Patricia Slepian, and Dr. Jean Straub of the University of Pennsylvania; Mr. David Fauser of the Springfield, Pennsylvania Junior High School; and Dr. George Kirk of the Smyrna, Delaware Public Schools assisted in trying out and refining tests.

Dr. David Fink of the University of Maine made available his tests of rigidity and helped in devising scoring methods for the various individual tests. Miss Estoy Reddin of the University of Pennsylvania assisted in the administration of individual tests.

Miss Arlene Krochmal, Miss Judy Manning, and Miss Virginia Maull of the University of Pennsylvania assisted in scoring group tests and in statistical work. Mr. Richard Heisler of the University of Pennsylvania helped in the construction of the apparatus test and in tabulating the scores on all tests.

Special thanks are due Dr. Ethel Maw of Bryn Mawr College who worked with us as research psychologist. In addition to helping with test selection, Dr. Maw assumed major responsibility for scoring the ideational fluency and similar non-objective tests and constructed the test of critical thinking which appears under her name in the Appendix.

Special thanks are also due Mrs. Irene Bottomley for her care in checking copy and preparing stencils.

CONTENTS

PREFACE ii

I INTRODUCTION 1

Problem solving. Selection of total sample. Selection of criterion tests. Selection of good and poor problem-solvers. Validation of selection. Administration of tests.

II DESCRIPTION OF TESTS 15

Achievement tests. Intelligence test. Tests of reasoning and problem solving. Tests of problem recognition. Word association tests. Memory test. Ideational fluency tests. Tests of closure. Tests of judgment or evaluation. Preferences test. Summary.

III DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS 33

Differences in problem recognition; word association; memory; ideational fluency; closure; judgment; preferences. Differences in variability. Sex differences. Differences in errors.

IV DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS (CON.) 48

Administration of tests. Einstellung tests. Color naming. Similarities. Disc-transfer problem. Persistence tests. Free association. Box problem. Experimentation. Problem-solving apparatus.

V IMPORTANCE OF CERTAIN INTELLECTUAL VARIABLES IN PROBLEM SOLVING 66

The predictor variables: problem recognition, word fluency, ideational fluency, closure, judgment. Results: problem recognition, word fluency, ideational fluency, closure, judgment, intelligence, reading.

VI SUMMARY AND CONCLUSIONS 73

Summary. Conclusions. Concluding remarks.

REFERENCES 83

APPENDIX 88

TABLES

1. Intercorrelations, Standard Deviations and Reliability Coefficients of Criterion Tests 8
2. Intelligence Quotients and Composite Measures of Problem-Solving Ability 10
3. Samples of Good and Poor Problem-Solvers by School, Grade and Sex 11
4. IQ Distributions of Good and Poor Problem-Solvers 11
5. Means of Good Problem-Solvers and Poor Problem-Solvers on Four Tests and Significance of Differences 12
6. Good and Poor Problem-Solvers as Ranked by Teachers 13
7. Mean Scores of Good and Poor Problem-Solvers on Tests of Problem Recognition and Significance of Differences 34
8. Mean Scores of Good and Poor Problem-Solvers on Word Association Tests and Significance of Differences 36
9. Mean Scores of Good and Poor Problem-Solvers on Tests of Ideational Fluency and Significance of Differences 39
10. Mean Scores of Good and Poor Problem-Solvers on Tests of Closure and Significance of Differences 40
11. Mean Scores of Good and Poor Problem-Solvers on Tests of Judgment and Significance of Differences 41
12. Number and Percentage of Good and Poor Problem-Solvers Successful on Disc Transfer and Median Number of Moves 54
13. Mean Number of Switches by Good and Poor Problem-Solvers on the Persistence Tests 56
14. Mean Scores of Good and Poor Problem-Solvers on Five Experimentation Exercises and Significance of Differences 60
15. Frequency of Errors by Good and Poor Problem-Solvers at Each Stage of First Apparatus Problem and Significance of Differences 63
16. Number and Percentage of Good and Poor Problem-Solvers Successful on the Problems of the Second Part of the Apparatus Experiment 64
17. Means and Standard Deviations of the Problem-Solving Criterion and the Tests for Seven Predictor Variables for 513 Subjects 68

18. Coefficients of Correlation Between the Variables 69

19. *Beta* Coefficients and Multiple Coefficients of Correlation Between Problem Solving and Selected Variables 70

20. Summary of Group Test Statistics for Good and Poor Problem Solvers 75

21. Summary of Results of Individual Tests of 50 Good and 48 Poor Problem-Solvers 77

22. Summary of Correlation Statistics of Criterion Measure of Problem-Solving Ability and Selected Predictor Variables for 513 Subjects 79

CHAPTER I

INTRODUCTION

This report presents a study of differences between good and poor problem-solvers in two suburban junior high schools and an analysis of problem-solving ability, as defined. Specifically it is concerned with the questions,

1. What observable variables discriminate between good and poor problem-solvers among junior high school pupils?
2. What is the importance of these variables in explaining or accounting for problem-solving ability?

Instead of Question 2, we had originally asked, "Are the discriminating variables different or of different importance at above average, average, and below average intelligence levels?" In a preliminary selection of good and poor problem-solvers, however, we found too few good problem-solvers at IQ levels below about 105 to permit reliable comparisons. (See Table 2) In our final selection of good and poor problem-solvers, we used deviations from the line of regression of problem-solving on IQ; hence, intelligence, as tested, is statistically controlled.

In this chapter we briefly discuss the term problem solving; the selection of the total sample; the selection of criterion tests; the selection of samples of good and poor problem-solvers and its validation; and the administration of the tests.

Problem Solving

It would be difficult, perhaps idle, to differentiate sharply between the terms thinking, problem solving, and reasoning. In his study of thinking, Humphrey (1951) brings the three together in the statement,

> Thinking as the term is understood in this book may be provisionally defined as what occurs in experience when an organism . . . meets, recognizes, and solves a problem. The term reasoning is preferred by some writers. (p. 311)

Humphrey goes on to say that, since problem and non-problem situations shade imperceptibly into each other, the term problem must necessarily be indefinite.

As used by some writers, perhaps the majority, thinking is the broadest of the three processes; reasoning the most restricted. Later in this chapter, in connection with the selection of criterion tests, we shall discuss in some detail several roles of reasoning in problem solving.

The term problem solving refers to a multiplicity of processes and events which have not yet been classified in any useful way. In his review, Corman (1958) finds so little agreement about what the term includes or should include that he considers "problem solving as meaning the behaviors that researchers, who say they are studying problem solving, study."

In a loose, general sense a problem exists every time an individual must choose between two or more alternatives. It is convenient and customary, however, to restrict the term problem to situations which for one reason or another appreciably hold up an individual in his effort to reach a goal. What the individual does in resolving the situation constitutes the processes which are subsumed under the phrase problem solving. These processes are not altogether mental. There are several theories and substantial evidence that problem solving has physiological counterparts.

There is some agreement that important among the mental processes are, (1) recognizing, preparing for, or orienting to the problem, (2) producing relevant materials, elaborating, and forming tentative solutions, and (3) judging, and selecting from, the alternatives that result. Johnson (1944, 1955) discusses these processes and emphasizes their interactions both with each other and with various mental abilities.

Studies of problem solving fall into two broad, partially overlapping classes: logical or philosophical and psychological. Dewey's (1910) well-known logical analysis of problem solving or reflective thinking has been influential for half a century. It provides a formal description of so-called scientific thinking or scientific method. It has been particularly attractive to educators doubtless because it elaborates an organized attack on problems which may be sharpened through training. It has stimulated many studies of the effect of instruction on problem-solving skills and scientific thinking. Essentially, Dewey's analysis presents the formal statement of the three processes listed above.

Psychological studies of problem solving or, more broadly, thinking have followed three main lines, the associational, the gestalt, and the psychometric. The three connect at various points, and there are, of course, differences within types as well as among types. The first is exemplified by the studies and theories of Thorndike, Guthrie, and Hull; the second by those of Wertheimer, Duncker, and Maier. Harlow (1951) gives a good account of the differences between these types in his discussion of the uni-process theorists or behaviorists and the dual-process theorists or gestaltists.

The psychometric studies and theories of problem solving and reasoning, of such factor analysts as Botzum, Adkins, and Guilford, are as much logical analyses as psychological. Johnson (1955) discusses in some detail the contributions of factor analysis to the psychology of thought and judgment. Here, we shall only remark that factor analysis, despite several theoretical limitations, is enormously useful in test selection and classification. In our selection of tests, we drew heavily from several factor analyses of reasoning ability.

Although according to the psychology of needs and affective states, every problem has as its source of energy certain tensions or needs, there has been little study of the role of motivation in problem-solving behavior. This has been due, perhaps, both to the tendency to consider problem solving as a purely intellectual process and the difficulty of reliably observing the motivation variables. Travers et al (1955) list among these variables a manipulatory, exploratory, or curiosity drive; an anxiety drive; an ambiguity intolerance drive; and achievement need. They believe,

> . . . In the case of complex problems it seems reasonable to postulate that reasoning variables and motivational variables will play a central role . . . In the domain of drives it seems likely that anxiety as a drive has greater relevance than an exploratory drive, and that achievement need has greater relevance than ambiguity tolerance. (p. 15)

Bruner et al (1956) build up a strong case for striking similarities between problem-solving strategies and scientific method, when motivation variables are disregarded, and thus connect with Dewey's logical analysis. Their study in part contradicts that of Burack (1950) who found that attacks on problems varied from individual to individual and from problem to problem.

There have been numerous, similar empirical studies of strategies and errors in thinking as revealed by analyses of what people do when trying to solve a problem. Bloom and Broder (1950) and Buswell (1956) are among the many investigators who contrast successful and unsuccessful attempts to solve problems; Brigham (1932) and Lawrence (1957) are among those who analyze errors and seek their explanation. These studies as a rule do not confirm any particular theory of problem solving, but they do frequently suggest ways in which problem-solving ability can be improved. To our knowledge, McNemar (1955) is the only investigator who has looked for differences between good and poor problem-solvers or reasoners, classified as such independently of the tasks in hand, in various abilities hypothetically related to problem solving. We return to these studies later.

Selection of Total Sample

For two reasons we decided to limit our study to seventh and eighth-grade pupils. First, such pupils are old enough for many of the tests that have been used in large-scale factor analyses of reasoning ability. Second, they have had little participation in the formal problem-solving activities which are ordinarily included in the mathematics, science, and social studies of the ninth and later grades. Questions about the effect of instruction on problem-solving ability are of course important, but our interest was in variables, relatively uncultivated by formal instruction, which differentiate good and poor problem-solvers.

Of the ten suburban junior high schools near Philadelphia we consulted in April, 1958, two were particularly eager to take part in the study, had essentially similar testing and guidance programs, and had excellent attendance records. (In extensive testing programs, the size of the working sample is of course greatly affected by attendance.) These were the Abington and Lansdowne-Aldan Junior High Schools. We included all of the seventh and eighth-grade sections in the former; and four seventh and five eighth-grade sections of the latter. In all, there were 316 seventh-grade pupils in 11 sections and 320 eighth-grade pupils in 12 sections.

The median chronological age in October, 1958, of the 636 pupils was 154 months with decile deviation from 143 to 165 months. The median intelligence quotient of the 615 available on the California Test of Mental

Maturity, Long Form, was 114 with decile deviation from 97 to 128. The distribution of IQ's is shown in the next chapter.

The sample appeared to be well representative of large-city suburban seventh and eighth-grade pupils. Actually, since the processes involved in problem solving appear to be common, the sample probably is fairly representative of other age and residence populations of similar IQ range.

Selection of Criterion Tests

There are no tests of problem-solving ability, as such; indeed, there would be little agreement about what such tests should contain. Guilford (1954) is of the opinion that there are numerous kinds of problem solving and for each kind of problem there are ordinarily several possible operations by which it can be solved. However, he believes that among the abilities most likely to be important in problem solving are general reasoning, fluency of ideas, flexibibility, evaluation, and originality. Travers et al (1955) suggest that variability-stereotypy, equivalence of reactions, reasoning, originality, learning set, and motivation variables "have relevance for the study of problem solving." Guetzkow (1951) postulates that problem-solving behavior may be accounted for, at least in substantial part, by the variables of set, fluency, and reasoning or solution-forming ability. McNemar (1955) believes that logical reasoning, as identified by Guilford and his associates, is "an indispensable, if not a fundamental, aspect of problem-solving."

There seems to be widespread agreement that reasoning ability is a crucial variable in problem solving, but whether there are different kinds of reasoning has not been resolved. Some factor analysts, perhaps the majority, have postulated and identified inductive and deductive reasoning factors. Holzinger and Swineford (1939), however, found that their battery of "thinking" tests had high loading on a general factor which appeared to be general reasoning ability. Their conclusion is essentially in agreement with Davis (1947) who identified in a battery of reasoning and practical judgment tests a general reasoning ability "probably closely akin to the principal mental function measured by ordinary intelligence tests."

In their partly experimental, partly theoretical analysis of the general reasoning factor, Guilford et al (1956b) report an attempt to separate the factor into such abilities as understanding and defining problems, thinking effectively by trial and error, and handling complex sets of elements and relations. The separation could not be satisfactorily achieved, and they conclude,

> . . . we may say it has been easier to decide what general reasoning is not than to say what it is. We have given reasons for believing it is not general intelligence, though it is undoubtedly one of the important intellectual abilities . . . it is not a generic ability to solve problems, though it undoubtedly plays a role in solving problems of certain types . . . By elimination and consistent indications of a positive nature, the best we can say is that general reasoning has something to do with comprehending or structuring problems of a certain kind in preparation for solving them. The range of problems that is pertinent is still to be determined. Whether this ability extends beyond problem solving is also to be determined. (p.172)

In sum, some investigators find that reasoning enters into the orientation process in problem solving; some, into the production or solution-forming process; others, into the judging process. It seems reasonable to assume that reasoning is the central ability elicited by the kinds of problems useful in controlled experimentation in the classroom and laboratory.

As our criterion tests, we selected five tests of reasoning and problem-solving ability: *Differential Aptitude Tests* of verbal, abstract, and mechanical reasoning; *Davis-Eells*(1) *Games*; and *Standard Progressive Matrices*. In addition, we constructed a test of reasoning, *Thought Problems*. The composition and mechanics of the tests and the score distributions are described in Chapter II; here, we need only to comment on our reasons for selecting them.

The *Differential Aptitude Tests* are well known and are among the best of their kind. In the excellent *Manual*, the authors state,

> The *Verbal Reasoning* test . . . is a measure of ability to understand concepts framed in words. It is aimed at the evaluation of the student's ability to abstract or generalize and to think constructively . . . (p.6)
>
> The *Abstract Reasoning* test is intended as a nonverbal measure of the student's reasoning ability . . . It involves the ability to perceive relationships in abstract figure patterns--generalization and eduction of principles from nonlanguage designs. (pp. 6-7)

> The Mechanical Reasoning test is essentially a new form of (Bennett's) Mechanical Comprehension Tests . . . The ability measured by the . . . test may be regarded as one aspect of intelligence, if intelligence is broadly defined . . . The score is affected by the previous experience of the subject but not to a degree that introduces serious difficulties in interpretation. (p. 7)

The tests sacrifice some factorial purity, but their predictive value has been found to be higher than purer tests. Although intended for Grades 8 through 12, the tests can be and have been used successfully in separating high and low groups in all junior high school grades, according to Wesman, one of the authors, in personal communication with us.

The Davis-Eells Games is a test of general intelligence or problem-solving ability. In the Manual, the authors state that the mental processes required by the test include,

> . . . the making of correct associations, the logical classifications of concepts, the discrimination of crucial differences, the definition of analogous relationships, the drawing of reasonable inferences, the organizing of the elements of a problem into a meaningful pattern, the gaining of insight, and the mental activity of foreseeing a solution. (p.2)

In a study involving some 1200 pupils in Grades 4, 5, and 6, Tate and Voss (1956) found, among other things, that the great majority of children enjoyed taking the test and that it measured abilities not well measured by conventional intelligence tests. Although the test is relatively easy for seventh and eighth-grade pupils, we included it because of the breadth of skills it elicits and because it is unusually interesting to take.

Progressive Matrices are, according to the Manual, a test of a person's capacity to apprehend meaningless figures, see the relations between them, conceive the nature of the figure completing each system of relations presented, and so demonstrate a systematic method of reasoning. The test is considered by its author, Raven, and by Spearman to be loaded in "g". In his study, Rimoldi (1948) identified space, perception, memory, and induction factors. Adkins and Lyerly (1952) found that Matrices B, C, and D had loadings in a factor they identified as hypothesis verification. The test turned out to be relatively easy for our sample.

We constructed Thought Problems, Parts I and II, copies of which are included in the Appendix, in answer to our need for a relatively hard test containing reasoning problems of various sorts. About half of the problems were devised by Burt (1919) for his study of the development of reasoning, and many of them have since been used in various factor analyses of reasoning tests and in studies of problem solving. We experimented with the problems during April, May, and September, 1958, trying both multiple-choice and open-response forms. The latter seemed to be measuring some things not measured by the former. It may be that finding answers is a different and more discriminating reasoning test than choosing between given answers, one of which is known to be correct. At any rate, the open-response form seemed best suited to our purposes. We assembled the problems in two parts, so that test reliability could be easily estimated.

Selection of Good and Poor Problem-Solvers

We had originally intended to use all of the tests discussed in the preceding section in selecting good and poor problem-solvers, or if the reader prefers, good and poor reasoners. The *Mechanical Reasoning* test, however, brought out differences between boys and girls large enough to disturb the sex balance in the final samples. The *Matrices* test seemed, after reflection, to be more useful as a validating than as a criterion measure.

In our final selection, we used the average of scores on the *Verbal* and *Abstract Reasoning* tests, ½(VR+AR); the scores on the *Davis-Eells Games*; and the scores on *Thought Problems*. The *Davis-Eells* means showed a difference of about 2 in favor of the boys, significant but not large enough to distort selection. The intercorrelations, standard deviations s, and reliability coefficients r_1 are shown in Table 1. The reliability of the first was estimated by the Kuder-Richardson formula (21); the reliabilities of the second and third from half-test coefficients, stepped up.

Table 1

Intercorrelations, Standard Deviations and Reliability Coefficients of Criterion Tests

Test	2	3	s	r_1
1. ½(VR+AR)	.517	.767	8.42	.905
2. Davis-Eells		.474	4.25	.648
3. Thought Problems			6.88	.855

In arriving at a composite, we tried out methods developed by Kelley (1927) and Horst (1936). The ratio of weights obtained for the three tests by Kelley's method was roughly 5:6:5 and by Horst's method 3:5:3:. Based on the inverses of the standard errors of measurement, the weights were approximately equal. As Gulliksen (1950) points out, a good case can be made for weighting inversely as the error of measurement where no criterion is available. We decided on the latter weights; hence, the composite or criterion measure of problem-solving ability is simply the sum of scores on M(VR+AR), *Davis-Eells*, and *Thought Problems*. The reliability coefficient of the composite, estimated by Kelley's (1927) method, is about .93.

As noted earlier, the attempt to control intelligence, as tested, experimentally was unsuccessful, and we had to resort to statistical control. The distribution of IQ's and Composite scores for the 554 pupils who had scores in both is shown in Table 2. The coefficient of correlation between the two is .645, and the equation of the line of regression of composite scores on IQ is C' = .886(IQ) - 18.05. The deviations of the Composite scores from their line of regression on IQ are theoretically independent of IQ. Pupils deviating most in the positive direction from the line have the greatest amount of problem-solving ability, as measured, independent of IQ; pupils deviating most in the negative direction from the line have the least. Actually, we worked from four tables similar to Table 2, one for each grade in each school. This was desirable to guarantee a balance of grades and schools in the final samples. (The slopes of the four regression lines were not significantly different from each other or from the line whose equation is given above, but the intercepts were.)

On the basis of positive and negative deviations of Composite scores from the lines of regression, we selected the samples of good and poor problem-solvers shown in Table 3. The samples are thus 117 good problem-solvers and 117 poor problem-solvers distributed as shown in the table. The mean of the former on the Composite was 98.7; the mean of the latter 66.6. The former deviated +14.9 on the average from the line of regression; the latter -15.1.

It is of interest to compare the IQ distributions of the two groups. These are shown in Table 4. The medians of the distributions are 114.8

Table 2

Intelligence Quotients and Composite Measures of Problem-Solving Ability

Intelligence Quotient

Problem Solving	70-	75-	80-	85-	90-	95-	100-	105-	110-	115-	120-	125-	130-	135-	140-	145-	130-	f
125-										1								1
120-											2	1	2			1		6
115-										3	1	2	2	1	1			10
110-									3		1	5	2	1	1			13
105-									1	3	4	5	5	2	2		1	23
100-					1	1		1	2	7	14	6	4			2		38
95-							1	3	6	13	8	3	4	2				40
90-							3	2	9	11	14	6	3	1	1			50
85-						5	1	11	12	16	9	4	2	1	1			62
80-					2	2	10	13	14	16	17	8						82
75-					4	2	8	8	9	9	5	4						49
70-				1	1	4	14	9	10	10	6		1					56
65-		2	1	2	2	1	9	10	7	7	3							44
60-				2	4	5	5	7	4	2	1							30
55-			1		1	4	5	3	4	1								19
50-	1	1			4	3	4	1		1								15
45-	1			2	4		2	1		1								11
40-					1		1											2
35-				1	1				1									3
f	2	3	2	8	25	27	63	69	82	101	85	44	25	8	6	3	1	554

and 113.6. (The median chronological ages of the groups were 156 and 154 months, respectively, in October, 1958.)

Table 3

Samples of Good and Poor Problem-Solvers by School, Grade and Sex

School	Grade	Good Problem-Solvers Boys	Good Problem-Solvers Girls	Poor Problem-Solvers Boys	Poor Problem-Solvers Girls
Abington	7	18	15	20	19
	8	16	18	16	16
Lansdowne	7	9	10	10	10
	8	15	16	17	9
	TOTAL	58	59	63	54

Table 4

IQ Distributions of Good and Poor Problem-Solvers

IQ	Good Problem-Solvers	Poor Problem-Solvers
145-		1
140-		2
135-		1
130-	7	4
125-	12	8
120-	17	19
115-	24	20
110-	23	19
105-	15	17
100-	8	14
95-	7	5
90-	4	7

Validation of Selection

There are neither infallible nor agreed upon criteria of problem-solving ability, and any attempt to validate a particular method of selecting good and poor problem-solvers is not unlike the blind leading the blind in a circle. Nevertheless, we believe several comparisons of the good and poor problem-solvers, as selected by the method described in the preceding section, are worth noting.

The mean grade equivalents of the two groups in the reading and arithmetic achievement tests, described in Chapter II, and the mean scores on the Mechanical Reasoning and Matrices tests are shown in Table 5. The reliabilities of the latter two tests, as estimated by Kuder-Richardson formula (21), were about .88 and .80, respectively; the reliabilities of the former two have generally been found to be satisfactory.

Table 5

Means of Good Problem-Solvers (GPS) and Poor Problem-Solvers (PPS) on Four Tests and Significance of Difference

Test	Mean GPS	Mean PPS	Mean Diff.	t Value of Diff.	d.f.	Significance level
Reading Achievement	9.56	8.27	1.29	6.93	232	.001
Arithmetic Achievement	9.08	7.96	1.12	6.40	217	.001
Mechanical Reasoning	33.7	22.5	11.2	7.36	232	.001
Matrices	47.8	38.0	9.8	10.1	231	.001

The differences in Table 5 are highly significant, the Matrices test in particular sharply discriminating between the groups.* The difference in arithmetic achievement would be expected, if only because arithmetic problems have usually been found to have heavy loading in the general reasoning factor. The difference in reading is interesting. Several factor analyses have found reasoning as well as verbal factors in reading tests. More than forty years ago, Thorndike observed,

> Understanding a paragraph is like solving a problem in mathematics. It consists in selecting the right elements of the situation and putting them together in the right relations . . . . The mind . . . must select, repress, soften, emphasize, correlate, and organize, all under the

---

*The Matrices test also brought out highly significant differences in the variability. In samples of this size the t test for means is little affected by differences in variability. However, such differences are important in their own right, and we return to them in Chapter III.

> influence of the right mental set . . . . .Reading an explanatory or argumentative paragraph . . . involves the same sort of organization and analytic action of ideas as occur in thinking of supposedly higher sorts.

At any rate, our good and poor problem-solvers as selected by four reasoning tests, two of which contain no reading and none of which contained difficult reading, differ markedly in reading ability.

Prior to the testing program, the mathematics teachers in the two schools had been asked to rank pupils according to the directions,

> Write "1" before the name of the pupil who seems to be most resourceful and ablest in working thought problems; "2" before the next ablest; "3" before the third ablest; and so on until 10 have been ranked.
>
> Write "30" before the name of the pupil who seems to be least resourceful and poorest in working thought problems; "29" before the next poorest; "28" before the third poorest; and so on until 10 have been ranked.

In Table 6, the ranks that had been so assigned by the teachers to our good and poor problem-solvers are shown. The chi-square test of independance, as applied to the table, resulted in a value of about 20 with 2 degrees of freedom. The corresponding probability is less than .001.

The five highly significant differences of tables 5 and 6 strongly confirm the belief that our method of selection did in fact result in 117 pupils who as a group are good at solving intellectual problems and in 117 who are relatively poor, with intelligence statistically controlled. How the groups differ in various simpler abilities will be considered in Chapters III and IV.

Table 6

Good and Poor Problem-Solvers as Ranked by Teachers

Group	Rank		
	1-10	None	21-30
Good	63	32	22
Poor	34	32	51

Administration of Tests

The first battery of tests discussed above, exclusive of the intelligence and achievement tests, was administered during the last two weeks of October, 1958. A second battery of group tests was administered during the second and third week of November, and two more group tests, Best Answer and Critical Thinking, were given in January. The various tests and score distributions are described in Chapter II. The individual tests, described in Chapter IV, were given during December, January, and February.

Outlines of the testing program are contained in memoranda to teachers, copies of which appear in the Appendix. One change in the program as described was made. The tests of the second battery were given to all pupils, instead of to only the selected good and poor problem-solvers, as originally planned. The change was necessary because of the administrative difficulty of separating pupils. The additional labor it entailed was more than offset by the resulting data, appropriate for the multivariate analysis discussed in Chapter V.

The details of the testing program were worked out by principals and guidance counselors. The first two batteries of tests were administered by the mathematics teachers at Abington and by the homeroom teachers at Lansdowne. So far as we could tell, all tests were well administered--unusually well administered in view of the extent of testing. Differences on tests between sections, between grades, and between boys and girls consistently appeared where they would be expected. Estimates of reliabilities of various tests were consistent with those reported by other investigators using the same tests in comparable samples. The names of principals and teachers who assisted with the program are included in the Preface.

CHAPTER II

DESCRIPTION OF TESTS

As noted in the preceding chapter, the testing program was divided into three parts. The first battery included tests of achievement and intelligence regularly given by the schools, and tests of reasoning and problem solving given under our direction during the last two weeks of October, 1958.

The second battery included group tests of recognition of problems, word association, memory, ideational fluency, closure, judgment, and preferences. All of the tests except those of closure and critical thinking were tried out and refined one or more times. All but two of the tests of the second battery were given during the second and third week of November. The third battery included individual tests of rigidity, word association, problem solving, and persistence. These tests were given during December, January, and February.

The first and second batteries are described in this chapter and the third in Chapter IV. The reliabilities of the tests of the second battery are reported in connection with the statistical comparisons of Chapter III.

Achievement Tests

Both Abington and Lansdowne regularly give achievement tests during the first month of the school year. Prior to 1958, both had given the *California Achievement Tests*, but Lansdowne changed to the *Iowa Every-Pupil tests of Basic Skills* after finding out that the grade equivalent scores in the two tests appeared to be comparable. Our own inspection of grade equivalents confirmed this, and we show below the combined distributions of reading and arithmetic achievement scores as obtained from the school records.

Arithmetic Achievement (N = 604)

Score	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0
	5.4	5.9	6.4	6.9	7.4	7.9	8.4	8.9	9.4	9.9	10.4	10.9	11.4	11.9	13.9
f	1	8	21	41	59	92	108	73	83	51	25	14	16	7	5

Reading Achievement (N = 628)

Score	4.5	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0
	5.4	5.9	6.4	6.9	7.4	7.9	8.4	8.9	9.4	9.9	10.4	10.9	11.4	11.9	12.9
f	7	8	14	21	37	70	91	89	79	81	69	32	18	5	7

Intelligence Test

Both Abington and Lansdowne regularly give the *California Test of Mental Maturity, Long Form* in the seventh grade during the first month of school. The intelligence quotients whose distribution is shown below were taken from the school records, those for the eighth-grade pupils dating back to September, 1957.

Intelligence Quotients (N = 615)

I.Q.	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150
	74	79	84	89	94	99	104	109	114	119	124	129	134	139	144	149	154
f	2	5	3	9	30	31	70	77	92	107	93	48	27	10	6	4	1

Tests of Reasoning and Problem Solving

The purpose of these tests was to provide criteria for the selection of good and poor problem solvers. All of the tests except *Thought Problems* are standardized. A copy of the latter is shown in the Appendix. We briefly describe the tests and give the score distributions below. Their use in constructing and validating a composite measure of problem-solving ability was discussed in Chapter I.

Verbal Reasoning, Form A, is one of the *Differential Aptitude Tests* of the Psychological Corporation. It consists of fifty verbal analogies. Each item consists of a sentence, incomplete at beginning and end, and the subject is asked to choose two words (each from four alternatives) which best complete the sentence. Directions for administering and scoring accompany the test.

Sample: (2) is to water as eat is to (C).

1. continue	2. drink	3. foot	4. girl
A. drive	B. enemy	C. food	D. industry

Verbal Reasoning (N = 605)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44
f	4	23	46	66	90	97	89	58	47	29	27	14	6	7	2

Abstract Reasoning, Form A, is another of the Differential Aptitude Tests of the Psychological Corporation. It contains fifty items consisting of rows of designs or figures. In each row, there is a trend discernible from the first to the fourth figure. The subject chooses from five alternatives the design or figure which should be next, or fifth, in the series according to the trend in that row. The tests are administered and scored according to the accompanying directions.

Abstract Reasoning (N = 605)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47
f	28	15	14	17	23	28	42	53	79	61	65	71	47	40	18	4

½(VR+AR) is merely the average of the scores on the above Verbal and Abstract Reasoning tests. Since the mean for the former is about 18 and for the latter about 26, ½(VR+AR) is weighted more heavily with abstract than with verbal reasoning.

½(VR+AR) (N = 605)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44
f	9	13	31	28	46	46	102	78	78	61	56	25	22	8	2

Mechanical Reasoning, Form A, another of the Differential Aptitude Tests, contains sixty-eight items consisting of pictures and questions about the pictures. The problems presented deal with such topics as centrifugal force, levers, pulleys, gears, volumes, weights. The test is administered and scored according to accompanying directions.

Mechanical Reasoning (N = 604)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53	59
f	11	7	9	21	30	54	48	57	52	52	60	36	47	32	28	25	20	11	4

Davis-Eells Games is a standardized test published by the World Book Company. It is composed of four kinds of items intended to evoke problem-solving skills. One set of items, the one dealing with money, is omitted in the present study. This shortens the test to fifty-four items. The shortened test is administered in one session of sixty minutes. In all other respects, standard procedures for administering and scoring are followed.

Davis-Eells Games (N = 610)

Score	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37	38 39	40 41	42 43	44 45	46 47	48 49	50 51	52 53
f	1	1	0	0	2	4	9	3	24	22	56	107	113	117	95	47	9

Standard Progressive Matrices, constructed by the British psychologist, J. S. Raven, is available in this country from the Psychological Corporation. It consists of five sets, A, B, C, D, and E, of twelve items each. In set A, geometrical patterns are presented in which a part is omitted, and the subject is asked to select from a number of alternatives the one which best completes the pattern. In sets C, D, E, each item consists of three rows and three columns of geometrical figures with a blank in the lower right cell. The figures in rows and columns are changing according to some rule, and the subject is asked to select from a number of alternatives the figure which completes row and column. Set B contains some items essentially similar to Set A and some essentially similar to Sets C, D, and E.

Standard Progressive Matrices (N = 613)

Score	6 8	9 11	12 14	15 17	18 20	21 23	24 26	27 29	30 32	33 35	36 38	39 41	42 44	45 47	48 50	51 53	54 56	57 59
f	2	2	2	2	2	6	12	5	27	38	46	61	119	106	108	55	19	1

Thought Problems is an original test composed of problems from several sources. Some of the items are from Burt's (1919) tests of reasoning; some are old puzzle problems of unknown origin; and some are adaptations and original problems devised especially for this test. The test is divided into two parts of twenty problems each. Sixty minutes are allowed for each part. The score is the number of problems solved.

Thought Problems (N = 592)

Score	2 3	4 5	6 7	8 9	10 11	12 13	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37
f	7	16	28	56	59	62	70	70	62	48	27	23	24	12	16	6	3	3

Tests of Problem Recognition

Recognition of, formulation of, or orientation to, the problem has consistently appeared in the literature on problem solving and is usually thought of as a crucial part of the preparation process. Its importance is supported by common sense as well as by research. One cannot seriously quarrel with the sense of the dictum, "a problem well stated is a problem half solved."

In various studies of reasoning and problem solving, involving different methods and materials, inadequate formulation of the problem has been found to be among the commonest kinds of errors. Lawrence (1957) summarizes several reports of this type of error and provides additional evidence of its prevalence. Unless the problem solver correctly understands the logical conditions of the problem, as Buswell (1956) points out, he cannot separate relevant from irrelevant facts or recognize the need for further facts.

To our knowledge, Guilford (1955) is the only factor analyst who has concerned himself with the ability or abilities involved in problem recognition. Both his "sensitivity to problems" and "conceptual foresight" factors would appear to be related to it.

The tests we selected for testing problem recognition were adapted from or suggested by Guilford's studies (1955, 1956a, 1957). The tests are described below along with methods of scoring and distribution of scores. Copies of the tests are included in the Appendix.

Seeing Problems I asks the subject to write down problems which might come about in connection with a pencil and with a candle. Four minutes are allowed for each task. The score is the total number of sensible problems listed.

Seeing Problems I (N = 592)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47
f	12	7	23	39	78	79	75	74	58	45	32	34	18	16	1	1

Seeing Problems II is derived from *Seeing Problems I*. The score is the number of independent, sensible problems listed. Problems such as "Keeping the wax off the table" and "Finding a way to catch the wax" are not considered independent; while "Making the candle stand straight" and "Keeping the light from going out" are considered independent.

Seeing Problems II (N = 592)

Score	0	2	4	6	8	10	12	14	16	18	20	22	24
	1	3	5	7	9	11	13	15	17	19	21	23	25
f	16	16	19	45	64	100	139	76	64	33	10	8	2

Missing Facts consists of thirty problems in arithmetic. Some of the problems cannot be solved because necessary information is missing. The subject is asked to list what fact or facts are needed where necessary information is lacking and to write nothing where all necessary information is given. Thirty-five minutes are allowed for the test. The score is the number of blanks correctly filled or left empty.

Missing Facts (N = 601)

Score	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31
f	2	3	5	6	18	22	53	52	69	87	88	72	72	40	11	1

Planning I consists of seven problems. The subject is asked to tell what information is needed before a solution can be reached. Twenty minutes are allowed for the test. The score is the number of facts listed.

Planning I (N = 599)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47
f	4	2	10	26	70	98	97	99	60	97	18	9	3	4	1	1

Planning II is derived from Planning I. The score is the number of facts listed which are useful in arriving at a solution of the problem.

Planning II (N = 599)

Score	0	2	4	6	8	10	12	14	16	18	20	22
	1	3	5	7	9	11	13	15	17	19	21	23
f	5	3	25	84	125	142	113	67	21	6	6	2

Word Association Tests

Whether or not one subscribes to association theory as an adequate explanation of thinking, there can be no serious doubt that association is an important process in thinking. Humphrey (1951) remarks,

> It may be said that the history of the psychology of thinking consists largely of an unsuccessful revolt against the doctrine of association. (p. 28)

Word association tests have an old and honorable history in the experimental study of association. Johnson (1955) believes that the word association experiment in its many forms yields considerable information on preparation for thought and bridges the gaps between simple reaction and the complex processes of problem solving. He summarizes the research since about 1940; Woodworth (1938) covers earlier research.

Beginning with Thurstone's (1938) monumental study, word association tests have been used in numerous factor analyses of mental abilities and have consistently yielded a factor which is currently identified as word fluency.

We selected four word-association or fluency tests similar to those used by McNemar (1955) in her study of good and poor reasoners. Descriptions of the tests, methods of scoring, and distributions of raw scores follow. Copies of the tests are included in the Appendix. It will be seen that the tests specify increasingly restricted conditions of word association.

Words is a test of word association with a minimum of restrictions. The subject is asked to write as many words as possible in three minutes. The score is the total number of recognizable words minus duplications. A retest was given after about a week to provide an estimate of reliability.

Words (N = 581)

Score	15	20	25	30	35	40	45	50	55	60	65	70	75	80
	19	24	29	34	39	44	49	54	59	64	69	74	79	84
f	1	2	10	27	52	77	119	102	70	57	30	20	9	5

First Letter, like Words, is a word association test, but it has the restriction that the words must begin with a given letter. There are two parts. In the first part, the subject is asked to write words beginning with P; in the second, words beginning with B. Three minutes are allowed for each part. The score is the total number of recognizable words on the two parts minus duplications.

First Letter (N = 583)

Score	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
	14	19	24	29	34	39	44	49	54	59	64	69	74	79	84	89	94	99	104
f	1	5	10	22	42	58	94	89	78	69	45	38	20	5	2	0	1	3	1

First and Last Letters is a test devised by Thurstone (1938), and is more restrictive than *First Letter*. In the two sections of the test, the subject is asked to write as many words as possible beginning with S and ending with L, and beginning with C and ending with T. Four minutes are allowed for each section. The score is the total number of correct words written.

First and Last Letters (N = 587)

Score	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39
f	10	16	47	57	82	101	81	64	43	32	24	17	7	2	2	1	0	0	1

Similar Words is an adaptation of Thurstone's (1938) *Controlled Association Test*. Two common words, large and fast are given, and the subject is asked to write words of the same or about the same meaning. Four minutes are allowed for each word. The score is the total number of correct words.

Similar Words (N = 594)

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
f	1	2	12	27	49	69	95	66	76	65	56	35	25	9	2	1	2	1	1

Memory Test

Since thinking is by definition a high-level response, it necessarily draws upon information in store and makes use of responses already established. Beyond this formal statement, however, little can be said about the role of memory in thinking.

In his review of the physiological evidence, Stellar (1957) hypothesizes that three memory mechanisms are operative; one for immediate memory, operative perhaps up to five minutes; a second for longer memories, on the order of hours or days; and a third, responsible for long-standing memories. Stellar's hypothesis may explain the indifferent correlation between memory span and delayed recall and account in part for varying notions about the role of memory in intelligence.

Guilford (1957) separates memory from intellect and believes that memory span, at least, has been overrated as a measure of intelligence. On the other hand Johnson (1955) believes that memory span may be crucial in thinking. He remarks,

> If our hypothesis about immediate memory span and thinking is correct, people with small spans would have difficulty in solving problems or making judgments that require attending to many objects, facts, or inferences at once. As compared to those who have large spans, they would be more liable to errors of omission, and would prefer conceptual systems of two or three divisions. (p. 92)

Bartlett (1958) is in essential agreement with Johnson. He concludes,

> The experiments . . . have shown that, so far as initial evidence goes, the most common feature of unsatisfactory thinking is that detail which is required to give the information its full directional properties is overlooked or forgotten. (p. 88)

Even less is known about delayed recall, whether it be over a few hours, days, or years. It may be that trying to remember how similar problems have been solved would interfere with appreciation of the demands of the new problem, i.e., it may be that long-term memory is inversely correlated with flexibility. Bartlett (1958) hints at this in the words,

> . . . it does seem that the claim of thinking . . . to be objectively appropriate, demands very great and literal loyalty to the immediate evidence, while equally its claims to be adaptable demand a capacity to treat the remote evidence as highly fluid in its characteristics. (p. 92)

We had not originally intended to include memory in our list of variables, and the memory test described below was added because one of our colleagues, Mr. Carl Hoffman, became interested in seeing whether various tests of immediate, delayed, associative, and incidental memory differentiated between the good and poor problem solvers in our sample of Lansdowne eighth-grade pupils. We shall return to the memory test and some of Mr. Hoffman's findings in a later section. A copy of the test is included in the Appendix.

Memory is a test of delayed recall. Two weeks after taking the *Thought Problems* tests, the subject is asked to list as many of the problems as can be remembered. It is not necessary to state the problem exactly. The test concludes with the questions, "Did you talk about any of the problems with someone outside of class?" and "Did you talk about any of the problems in class?" Fifteen minutes are allowed for the test. The score is the number of problems recalled.

Memory (N = 568)

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
f	44	74	72	78	60	68	45	39	22	24	9	14	8	6	2	1	1	0	1

Ideational Fluency Tests

At least five factor analyses of mental abilities--Carroll (1941), Bechtoldt (1947), Taylor (1947), Adkins and Lyerly (1952), and Guilford et al (1952)--have identified a factor, ideational fluency. The factor is described as the facility to call up ideas when quantity rather than quality is emphasized and is elicited by tests requiring the writing down of ideas about a given topic as fast as possible. It has generally been assumed that the variance of the score depends on the number of ideas rather than any aspect of expressing them.

Ideational fluency is a relatively new construct, and there has been little study of its correlatives and predictive value. At present, the case for it in thinking is more logical than empirical. Certainly, it is logical to suppose that the person who is capable of rapidly producing ideas, other things equal, will solve more problems than one less capable.

Our preliminary work with the ideational fluency tests described below suggested that different topics or tasks tended to produce different results and that the number of ideas independent of quality was not as pertinent to our study as certain refined scores. Hence, we retained all of the tests we had experimented with and took refined scores as well as gross scores from the tests. In the next chapter, we consider the question of what abilities the refined scores reflect.

Descriptions of the tests, methods of scoring, and distributions of scores are given below. Copies of the tests are included in the Appendix.

Reasons I is similar to a test used by Adkins and Lyerly (1952). It consists of two questions, "Why do people like to have trees in their back yards?" and "Why do people want to live in cities?" The subject is asked to write as many reasons as possible. The score is the number of different, sensible reasons given for both questions. Six minutes are allowed for each question.

Reasons I (N = 598)

Score	3	6	9	12	15	18	21	24	27	30	33	36	39	42	. . .	63
	5	8	11	14	17	20	23	26	29	32	35	38	41	44	. . .	65
f	5	18	54	77	82	84	88	66	34	24	28	11	16	10	. . .	1

Reasons II is derived from Reasons I. It is the total number of independent, sensible reasons given. For example, "To protect them from the sun" is not considered to be independent of "To provide shade." The replies "For beauty," and "For fruit" are considered to be independent.

Reasons II (N = 598)

Score	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32
	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33
f	2	8	35	57	83	93	87	74	65	44	26	14	3	5	1	1

Ideas I is similar to Topics used by Adkins and Lyerly (1952) which they credit to R. B. Cattell. The subject is instructed to write as many ideas as possible about two topics, "A man going up a ladder" and "A man driving a truck down a street." Five minutes are allowed for each topic. The score is the number of ideas listed.

Ideas I (N = 602)

Score	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	...	63
	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53	...	65
f	1	5	7	22	43	45	65	54	52	62	63	52	43	24	28	31	4	...	1

Ideas II is derived from Ideas I. It is the number of independent ideas listed for the two topics. For example, "Why is he going up?" and "Is he going to paint the roof?" are not considered to be independent. "The man is wearing a red shirt" and "The man will fall" are considered to be independent ideas.

Ideas II (N = 602)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44
f	12	13	16	30	66	76	97	90	58	54	51	22	10	5	2

Groups of Things is similar to the test used by Adkins and Lyerly (1952) which they credit to C. W. Taylor. The subject is asked to list the names of things that are round or could be called round and of things that are like a square or rectangle. Three minutes are allowed for each task. The score is the number of things listed which meet the specifications.

Groups of Things (N = 591)

Score	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
	8	11	14	17	20	23	26	29	32	35	38	41	44	47	50	53
f	3	12	28	59	61	92	98	69	56	45	27	21	8	7	3	2

Consequences I is similar to a test used by Guilford et al (1956). It asks the questions, "What would happen if everyone suddenly went blind?" and "What would happen if all laws were suddenly wiped out?" The subject is asked to write as many consequences as possible. Eight minutes are allowed for each task. The score is the number of immediate consequences listed.

Consequences I (N = 600)

Score	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
	4	9	14	19	24	29	34	39	44	49	54	59	64	69	74	79	84	89
f	10	53	90	90	97	67	59	43	36	35	14	2	0	2	1	0	0	1

Consequences II is derived from Consequences I. It is the number of remote or far-reaching consequences listed.

Consequences II (N = 600)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	. . .	63	66
	2	5	8	11	14	17	20	23	26	29	32	35	38	. . .	65	68
f	220	125	63	49	33	38	18	15	14	10	4	2	7	. . .	1	1

Uses I is similar to a test used by Guilford et al (1956). Two common objects, newspaper and brick, are given, and the subject is asked to list as many uses for each as he can. Four minutes are allowed for each task. The score is the total number of uses listed.

Uses I (N = 595)

Score	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45
	2	5	8	11	14	17	20	23	26	29	32	35	38	41	44	47
f	5	11	38	70	79	83	84	69	52	37	23	21	13	6	2	2

Uses II is derived from Uses I. It is the number of independent uses listed. Uses such as "To swat flies" and "To kill insects" are not considered to be independent.

Uses II (N = 595)

Score	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
f	8	21	35	66	83	73	75	73	58	40	33	13	7	7	1	1	0	1

Tests of Closure

The factors, flexibility of closure and speed of closure, were apparently first identified by Thurstone (1944, 1949) in his study of perception and mechanical aptitude. The former may be defined as the ability to identify a definite figure in the midst of or in spite of perceptual distractions; the latter, as the ability to unify an incomplete or discrete perceptual field into a single percept. Thurstone's study showed that the closure factors were related to a reasoning factor, induction.

Botzum (1951), Adkins and Lyerly (1952), and Pemberton (1952) identified both factors in their analyses of reasoning tests. Whether or not the factors are related is a moot question. Botzum reports substantial positive loadings for both on a second-order configurational or closure factor; Pemberton reports that the two do not have loadings on the same second-order factor; Matin and Adkins (1954) report a positive loading for flexibility of closure and a negative loading for speed of closure on a second-order factor tentatively identified as flexibility in analysis.

There is no disagreement, however, about relationship between closure and reasoning factors. As Johnson (1955) points out, this relationship adds some support to the arguments of Binet and the gestalt psychologists for a connection between perceptual reorganization and cognitive reorganization.

The closure tests we selected are included in the Appendix and are briefly described below.

Mutilated Pictures was adapted from Thurstone's test and made available to us by Adkins and Lyerly. Each item of the test contains two pictures alike in every respect except that in one of the pictures some part has been erased. The subject is asked to mark the picture in each pair that has a part missing. Ten minutes are allowed for the test. The score is the number of items marked correctly minus the number marked incorrectly.

Mutilated Pictures (N = 596)

Score	6 7	8 9	10 11	12 13	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35
f	1	1	1	1	3	5	10	10	19	47	40	52	80	118	208

Hidden Pictures was made available to us through the courtesy of Botzum. It consists of six pictures within which human faces and figures can be found embedded in the total picture. The subject is told how many persons to look for in each picture. Twelve minutes are allowed for the test. The score is the number of persons found.

Hidden Pictures (N = 597)

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f	1	7	14	38	68	77	86	101	71	63	29	24	10	3	5

Incomplete Pictures is Thurstone's adaptation of the Street Gestalt Completion Test. The subject is asked to identify an object or objects from the outline or parts given. The score is the number of objects correctly identified.

Incomplete Pictures (N = 600)

Score	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
f	2	1	4	13	7	15	20	34	31	39	43	54	68	58	62	65	45	22	15	2

Incomplete Words is Thurstone's Mutilated Words test. The subject is asked to identify words from the outline or parts given. Five minutes are allowed for the test. The score is the number of words correctly identified.

Incomplete Words (N = 599)

Score	4 5	6 7	8 9	10 11	12 13	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37
f	1	8	25	43	56	64	87	78	60	54	54	26	25	11	4	2	1

Concealed Figures is Thurstone's adaptation of the Gottschaldt Figures Test and was made available to us through the courtesy of Botzum. The subject is asked to identify simple designs hidden in more complex designs. The score is the number of correct identifications.

Concealed Figures (N = 589)

Score	15 17	18 20	21 23	24 26	27 29	30 32	33 35	36 38	39 41	42 44	45 47	48 50	51 53	54 56	57 59	60 62	63 65	66 68
f	3	4	5	7	16	23	35	64	57	63	69	66	62	50	31	24	6	4

Tests of Judgment or Evaluation

Johnson (1955) summarizes the role of judgment in problem solving in the words,

> When the problem to be solved is a complex one that leads the thinker through a variety of intellectual activities, the last phase of the problem-solving episode is usually a judgmental or decisive process. The thinker halts his productive activity to judge the merit of what he has produced . . . . The term judgment, although it has had many meanings throughout its long and honorable career, serves very well at the present moment in the history as a name for this conclusive activity. (p. 280)
>
> In general, judgment is a conclusive or decisive process, not a productive one, that brings a thoughtful episode to an end. In the case of highly differentiated human beings it is a retroflex process in that the thinker takes into account the motivational and instructional conditions that initiated the thoughtful episode. (p.282)

Although judgment would seem to be involved one way or another at every point in problem solving, it probably is true that it does not come fully into play until the problem solver has reached possible solutions. At any rate, it is convenient to treat it as the final process in problem solving.

Guilford and his associates (1956a, 1957) have attempted to explore the factorial composition of judgment or evaluation. Their studies suggest that among the abilities involved are practical estimation, practical judgment and logical reasoning.

As would be expected, tests of such abilities are lacking in factorial purity. For example, in a sample of 758 college students, the five subtests--Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments--of the Watson-Glaser test of critical thinking (largely logical reasoning) have intercorrelations whose median coefficient is about .30. When the coefficients are corrected for attenuation, the median is about .50. The coefficients of correlation of the total tests with various intelligence tests tend to cluster around .45.

We constructed two relatively complex tests of judgment: Estimation and Best Answer. These were tried out and refined twice during April, May, and September, 1958. A third test, Critical Thinking, was constructed especially for our study by Dr. Ethel Maw.

The tests, methods of scoring, and score distribution are described below. Copies of the tests are included in the Appendix.

Estimation consists of thirty multiple-choice items involving size or number. In scoring, two points are given for the best answer to an item and one point is given for an answer adjacent to the best answer. The score is the total number of points. Thirty minutes are allowed for the test.

Estimation (N = 600)

Score	12 13	14 15	16 17	18 19	20 21	22 23	24 25	26 27	28 29	30 31	32 33	34 35	36 37	38 39	40 41	42 43	44 45	46 47
f	1	6	10	30	37	56	70	66	97	73	75	36	25	8	6	1	2	1

Best Answer was suggested by Cardall's (1950) *Test of Practical Judgment*. It contains thirty-six questions, each followed by four possible answers. The subject is asked to choose the best answer and the next-best answer for each question. In scoring, the answers are ranked from 1, best, to 4, worst, on a key. The differences between the subject's rankings and the key rankings are summed and subtracted from 100. Hence, the highest possible score is 100, and the lowest -44. Twenty-five minutes are allowed for the test.

Best Answer (N = 578)

Score	9 11	12 14	15 17	18 20	21 23	24 26	27 29	30 32	33 35	36 38	39 41	42 44	45 47	48 50	51 53	54 56	57 59	60 62	63 65	66 68	69 71	72 74	75 77
f	1	1	1	2	5	5	8	18	18	27	30	40	57	77	60	68	60	42	32	19	5	1	1

Critical Thinking consists of a number of paragraphs containing information about topics in general science and social studies. Following each paragraph is a list of statements relating to the paragraphs. The subject is asked to mark each statement as true, probably true, false, probably false, or indeterminate in light of the information given. Thirty-five minutes are allowed for the test. The score is the number of statements correctly marked.

Critical Thinking (N = 560)

Score	15 17	18 20	21 23	24 26	27 29	30 32	33 35	36 38	39 41	42 44	45 47	48 50	51 53	54 56	57 59
f	6	14	23	42	70	87	84	97	53	35	27	11	7	3	1

Preferences Test

To our knowledge, no one has studied the relationship between liking to think and problem-solving ability, although several interest inventories include categories which might be loosely described as preferences for thinking activities and one, the *Personality Research Inventory* (Saunders, 1955), includes explicitly a liking-to-think category.

It would seem reasonable to hypothesize that preference for thinking activities would discriminate between good and poor problem-solvers. Accordingly, we experimented with tests of preferences during April, May, and September, 1958, and developed the test described below. A copy of the test is included in the Appendix.

Preferences consists of forty-three pairs of activities or options. The subject is asked to check the one in each pair he would prefer if a choice had to be made. Twenty minutes are allowed. The score is the number of items in which the subject chooses the activity which would ordinarily require the greater amount of thinking.

Preferences (N = 585)

Score	3 5	6 8	9 11	12 14	15 17	18 20	21 23	24 26	27 29	30 32	33 35	36 38	39 41	42 44
f	1	7	38	61	67	91	75	88	74	38	27	13	4	1

Summary

In summary, the first battery of tests included the *California Test of Mental Maturity, Long Form*; the *California* or *Iowa Every-Pupil* achievement tests of reading and arithmetic; the *Differential Aptitude* tests of verbal, abstract, and mechanical reasoning; *Davis-Eells Games*; *Standard Progressive Matrices*; and a reasoning test, *Thought Problems*. The results of the first three were available from school records; the latter tests were given under our direction by teachers. The uses of the first battery of tests in constructing and validating a composite measure of problem-solving ability were discussed in Chapter I.

The second battery included three tests of problem recognition; four word association tests; a memory test; five tests of ideational fluency; five tests of closure; three tests of judgment, and an inventory of preferences for thinking activities.* These tests were given by classroom teachers under our direction. Their power to discriminate between good and poor problem-solvers is discussed in the following chapter.

---

*We had originally hoped to find or construct relatively "pure" group tests of abstraction and generalization, location of error, and difficulty-reduction tendency, but were unable to do so. The difficulty-reduction tendency, in particular, has been and remains intriguing to us, The tendency of people to work problems easier than those posed has been remarked by numerous students of problem-solving behavior. It is reasonable to suppose that good and poor problem-solvers would differ greatly in the tendency.

CHAPTER III

DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS

The tests of problem recognition, word association, memory, ideational fluency, closure, judgment, and preferences included in the second battery were discussed in the preceding chapter, and the distributions of the scores of all pupils taking the tests were given. The means of the good and poor problem-solvers on the various tests are compared in this chapter.

Before proceeding to the comparisons, we need to note several limitations. First, since different pupils were absent during different tests, it was impossible to limit comparisons to identical pupils in the two groups without substantial loss of sample size. However, at least about 94 per cent of both groups were present for the great majority of tests, and in no test were fewer than 105 of either group present. The number involved in any comparison is of course two more than the number of degrees of freedom shown in tables following.

Second, the score distributions on several tests departed from normality to an extent inconsistent with assumption of population normality. Moreover, the assumption of equality of variance was frequently questionable. Where the assumptions supporting the t-test were especially untenable, we applied Wilcoxon's rank-sum test. The normal deviates $\underline{z}$ from the rank-sum test are shown at the foot of the tables. Actually, in samples of the present size, both assumptions of normality and variance equality can be safely relaxed. Apart from assumptions, however, heterogeneity of variance raises questions which we shall consider after discussing significance of differences between means.

Third, several of the tests, particularly those of ideational fluency, were difficult to score. The harder of these were scored independently by two scorers. Although individual scores assigned by the two frequently differed by several points, in only one test were differences between means materially affected. For that test, *Seeing Problems* II, a third scorer reconciled the discrepancies. In all cases the principal scorer did not know whether the examinee was a good or a poor problem-solver.

Finally, there are significant differences between boys and girls on several tests--the word association tests in particular. The means of good and poor problem-solvers on these tests are affected disproportionately, since there are in all 58 boys and 59 girls in the former, and 63 boys and 54 girls in the latter. However, the disproportion is slight, and adjustment of means for sex differences would not change conclusions. We discuss sex differences further in a later section.

Differences in Problem Recognition

The statistics from the tests of problem recognition and test reliabilities are shown in Table 7. All of the tests except *Seeing Problems I* bring out highly significant differences between good and poor problem-solvers. *Seeing Problems I*, as scored, is perhaps more a test of ideational fluency or free play of thought than of problem recognition. It will be noted that *Planning II* discriminates much more sharply than *Planning I* between the groups. The two tests differed only in method of scoring. In the latter, every fact not clearly irrelevant to the problem

Table 7

Mean Scores of Good and Poor Problem-Solvers on Tests
of Problem Recognition and Significance of Differences

Test	Reli-ability	Mean		Mean Diff.	t value of Diff.	d.f.	Level of Sig.
		GPS	PPS				
Seeing Problems I	.85[a]	19.55	19.69	-.14	-.16	217	.80
Seeing Problems II	.80[a]	12.91	10.46	2.45	4.24	217	.001
Missing Facts	.86[a]	21.76	15.65	6.11	9.95[b]	228	.001
Planning I	.90[a]	21.64	18.38	3.26	3.84	222	.001
Planning II	.64[a]	11.76	9.01	2.75	6.81[c]	222	.001

[a] Estimated from half-test scores in a 20 per cent random sample of total group of about 600.

[b] z Value from rank-sum test, 8.54.

[c] z Value from rank-sum test, 6.28.

was counted; in the former, only those facts which ordinarily would be useful in reaching a solution or decision. Here, as in the word-association and fluency tests discussed later, the more restrictive the requirements or scoring of a test, the greater the discrimination between groups.

We called these tests problem-recognition tests for want of a better name. They are relatively complex tests and elicit several abilities, but the multiple regression analysis of Chapter V shows that they make independent contribution to the criterion or composite measure of problem-solving ability. Partly out of curiosity, we tested for significance of differences between seventh-grade good and poor problem-solvers in the Missing Facts test with control on reading. We obtained an $F_{1,104}$ ratio of 14.8 significant at better than the .001 level.

All of the tests seem to be more or less concerned with recognizing or orienting to problems. Perhaps the title "appreciation of problems and their demands" would be more appropriate than "recognition of problems." Guilford's (1957) conceptual-foresight factor may be as good a name as any for the ability the tests primarily seem to elicit. Whatever the abilities are, the good and poor problem-solvers, as groups, have them in significantly different amounts.

Differences in Word Association

The word association tests of the second battery are similar to those used by McNemar (1955), and the results are comparable to hers.*

---

*We made no attempt to control speed of writing on the tests. McNemar found that her poor reasoners wrote significantly faster than her good and noted, "It is hard to believe that the High Group cannot write as fast as the Low; it seems more likely that the difference found reflects a conative difference--lesser willingness on the part of the Highs to compromise quality for the sake of quantity." We found evidence of sacrifice of quality for quantity in a good many hard-to-decipher words, but the frequency of such words was about the same in both groups. It is extremely doubtful that speed of writing affected mean differences between groups.

The t values corresponding to the mean differences on the four tests are .88, 2.33, 2.75, and 3.55, respectively, with about 220 degrees freedom, as shown in Table 8.

Table 8

Mean Scores of Good and Poor Problem-Solvers on Word Association Tests and Significance of Differences

Test	Reliability	Mean GPS	Mean PPS	Mean Diff.	t value of Diff.	d.f.	Level of Sig.
Words	.73[a]	51.80	50.50	1.30	.88	220	.30
First Letter	.86[b]	50.31	46.46	3.85	2.33	217	.05
First & Last Letter	.68[b]	14.39	12.62	1.77	2.75[c]	220	.01
Similar Words	.28[b]	8.22	6.90	1.32	3.55	221	.001

[a] Estimated from test-retest scores in a 20 per cent random sample of total group of about 600.
[b] Similarly estimated from half-test scores.
[c] z value from rank-sum test, 2.71.

McNemar's t values are .51, 2.50, 3.61, and 5.60 with about 90 degrees freedom. In both cases, the significance of differences increases with degree of restriction on association. The similarity is striking, particularly since McNemar's samples of "individuals with high and low reasoning ability" were mainly college sophomores.

Similar Words was too hard for our sample and shows very poor reliability; however, it is reasonable to assume that a more reliable test would, if anything, yield a greater mean difference. The fact that the significance of differences between groups increases as word association becomes more restricted indicates that, in the mental activity of free and controlled word association, good problem solvers differ from poor chiefly in the latter. McNemar observes,

> The fact that differentiation between good and poor reasoners increases with the degree of restriction placed on the response strongly suggests that the important variable, at least at the simple word level, is speed of selection of appropriate responses. It would be worthwhile to learn whether similar results could be obtained with less simple media than isolated words, say with ideas, plans of action, concepts, and hypotheses.

McNemar's observation is a summary rather than an explanation of findings. Just how the abilities elicited by word association tests relate to problem solving is a difficult question. Johnson (1955) offers several possible answers,

> . . . The more restricted the task, the easier it is to classify the response word as wrong and the more the task takes on the character of a problem . . . If the task is difficult, the subject is likely to take a critical attitude toward the words he has produced; a retroflex activity of judgment occurs. The subject turns around to verify the agreement between the solution and the requirements of the task. Thus we have in miniature a complete act of thought . . . (p. 151)
>
> If the schema is accepted as a logical concept which probably corresponds in some ways with psychological operations, a broadened account of its role in word association may be attempted. From the whole environment, perceived as a whole or in parts, at present or in the past, the thinker abstracts certain aspects connected in some way perhaps accidentally, with the present task . . . The thinker combines these influences according to his own abilities and biases into a schema. With this schema as a guide or model he searches his memory for a word that fits; the search model, then, is the conscious aspect of the preparatory set.
>
> In this account, or any other reasonable account of the role of the search model, it appears that the search model, which is conscious, guides associations, facilitating some responses and inhibiting others before they reach the threshold of consciousness. (pp. 154-155)

Johnson's conjectures are interesting. If there is in fact a mechanism that facilitates some responses while inhibiting others in accordance with the demands of the word-association problem, here at least the doctrine of association is not inconsistent with the directed nature of thought.

Differences in Memory

The mean number of thought problems recalled by the good and poor problem-solvers were 5.19 and 3.22, respectively. The t value of the difference was 5.12 with 218 degrees freedom. (The z value from the rank-sum test was 4.94.) The corresponding probability is less than .001.

The memory test included questions regarding whether the subject had talked about the problems in or outside class. Eighty-one good and 80 poor problem-solvers answered both questions. Of the former, 81 per cent had talked about the problems outside of class and 42 per cent in class, compared to 55 per cent and 38 per cent, respectively, of the latter. The difference between percentages talking about the problems outside of class is highly significant. Because of this difference and because the coefficient of correlation between memory scores and number of problems solved was about .59, we hesitate to draw any conclusions at all from the memory test. However, the difference in percentages who talked about the problems outside of class suggests that good problem-solvers are more interested in or concerned about the kinds of problems in the _Thought Problems_ test.

One of our colleagues, Hoffman (1959), is studying the relation of memory abilities to problem-solving ability in our Lansdowne eighth-grade samples. Of his various tests, only Saugstad's incidental-memory tests, a two-weeks delayed-recall test, and the Stanford-Binet sentence memory test differentiated the two groups significantly. The difference in one associative-memory test approached significance. In those tests, the good problem-solvers exceeded the poor.

Differences in Ideational Fluency

The statistics from the ideational fluency test are shown in Table 9. As noted in the preceding chapter all of the II tests, except _Consequences II_, are obtained from the I tests through more restricted scoring. The score on _Consequences I_ is the number of immediate consequences listed; on _Consequences II_, the number of remote consequences.

It will be noted that none of the I test means, except Consequences I, are significantly different, and that the difference on that test favors the poor problem-solvers. All of the II test means, except _Reasons II_, show significant differences in favor of the good problem-solvers.

The only restriction on quality in these ideational-fluency tests, as administered, is the quality level the subject deems satisfactory as he interprets the task to be performed. It seems to us that the most reasonable explanation is the somewhat vague and circular one that the good problem-solvers, as a group, are higher-quality or more self-critical performers. Certainly, their responses are deeper or more thoughtful than those of the poor problem-solvers.

Table 9

Mean Scores of Good and Poor Problem Solvers on Tests of Ideational Fluency and Significance of Differences

Test	Reli-ability	Mean GPS	Mean PPS	Mean Diff.	t value of Diff.	d.f.	Level of Sig.
Reasons I	.81[a]	19.61	20.04	-.43	-.51	225	.60
Reasons II	.71[a]	14.24	13.84	.40	.57	225	.50
Ideas I	.85[a]	29.84	29.15	.69	.65	226	.50
Ideas II	.83[a]	22.33	18.89	3.44	3.34	226	.001
Groups of Things	.66[a]	25.75	23.88	1.87	1.79	218	.10
Consequences I	.73[a]	20.79	26.19	-5.40	-3.35	220	.001
Consequences II	.74[a]	9.15	5.23	3.92	3.84[b]	220	.001
Uses I	.76[a]	17.61	19.03	-1.42	-1.29	221	.20
Uses II	.68[a]	13.23	11.18	2.05	2.69	221	.01

[a] Estimated from half-test scores in a 20 per cent random sample of total group of about 600.

[b] z Value from rank-sum test, 4.20.

The Consequences tests seem to confirm the interpretation. If the means of the two tests are combined, it will be seen that the poor problem-solvers excel in the total number of consequences listed. In Consequences I, where the score is the number of immediate consequences listed, they excel significantly. However, in Consequences II, where the score is the number of remote or far-reaching consequences, the good problem-solvers are significantly superior. The Planning tests discussed earlier in this chapter, also confirm the interpretation, and the results of the word-association tests are not inconsistent with it.

The results of the ideational-fluency tests may partly answer McNemar's (1955) implied question as to whether "degree of restriction placed on . . . less simple media than words, say (on) ideas, plans of action, concepts, and hypotheses" would differentiate good and poor reasoners. However, it may well be that tests in which the restrictions appear in the requirements will give different results than tests where the restrictions are enforced in the scoring. At any rate, the good and poor problem-solvers differ significantly in quality of response as elicited by our ideational-fluency tests.

Differences in Closure

Interest in the relation of closure to reasoning or problem solving stems, of course, from questions regarding the connection between perceptual and cognitive processes. Factor analysts have consistently found a speed-of-closure or a flexibility-of-closure factor, or both, in reasoning tests.

Of the closure tests whose statistics are shown in Table 10, the second, third, and fourth are concerned with flexibility of closure or the ability to unify an incomplete perceptual field into a single percept, while the last is concerned with flexibility of closure or the ability to identify a definite figure in spite of perceptual distractions. Where the first, Mutilated Pictures, belongs is not clear. Although Adkins and Lyerly (1952) found a relatively high loading for it on speed of closure, we are inclined to believe that, unless highly speeded, it has little in common with other closure tests. Wechsler believes his similar test, Picture Completion, to be a test of the ability to differentiate essential from unessential details.

Table 10

Mean Scores of Good and Poor Problem-Solvers on Tests of Closure and Significance of Differences

Test	Reliability	Mean GPS	Mean PPS	Mean Diff.	t Value of Diff.	d.f.	Level of Sig.
Mutilated Pictures	.86[a]	31.94	29.67	2.27	3.54[b]	221	.001
Hidden Pictures	.40[a]	6.96	6.42	.54	1.50	220	.20
Incomplete Pictures	.71[a]	17.39	15.97	1.42	2.90	221	.01
Incomplete Words	.88[a]	19.41	17.96	1.45	1.89	221	.10
Concealed Figures	.88[a]	46.46	44.12	2.34	1.88	224	.10

[a] Estimated from half-test scores in a 20 per cent random sample of total group of about 600.

[b] z Value from rank-sum test, 3.83.

We did not try out the closure tests, and our time allowances were not altogether satisfactory. The twelve minutes allowed for Hidden Pictures were not enough, and the test showed poor reliability. The time allowances on the other tests may have been too generous. Hence, we question whether the tests, as administered, are comparable to those of the factor analysts.

Of the closure tests, only Incomplete Pictures differentiate significantly between the two groups, although all approach significance. It seems probable that good and poor problem solvers differ significantly in closure abilities, but that the difference is relatively unimportant. This is confirmed by the multiple-regression coefficients reported in Chapter V.

The results from Mulilated Pictures, whatever their relation to closure, indicate that good problem-solvers are significantly superior in discrimination of detail, as measured.

Differences in Judgment

Of the various tests of the second battery, the judgment tests as a group bring out the greatest differences between good and poor problem-solvers. The t values and significance levels corresponding to the mean differences are shown in Table 11. In Chapter V, it will be seen that

Table 11

Mean Scores of Good and Poor Problem-Solvers on Tests of Judgment and Significance of Differences

Test	Reliability	Mean GPS	Mean PPS	Mean Diff.	t Value of Diff.	d.f.	Level of Sig.
Estimation	.37[a]	28.84	27.10	1.74	2.31	218	.05
Best Answer	.77[a]	55.31	45.04	10.27	7.70[b]	217	.001
Critical Thinking	.66[a]	38.17	29.92	8.25	9.01[c]	209	.001

[a] Estimated from half-test scores in a 20 per cent random sample of total group of about 600.

[b] z Value from rank-sum test, 6.64.

[c] z Value from rank-sum test, 8.86.

the composite score on these tests contribute nearly as much as reading and IQ to the variance of problem solving.

Estimation is the poorest of the three, both in reliability and discrimination between groups. On the basis of two tryouts* and refinements, we had expected it to work better. The present form is somewhat less reliable than the form it was refined from. Better reliability would add to its discrimination.

All three tests are relatively complex and measure different abilities. The coefficients of correlation of Estimation and Best Answer, Estimation and Critical Thinking, and Best Answer and Critical Thinking, based on the some 460 pupils who had scores in all three, were .28, .22, and .45, respectively. Corrected for attenuation, the coefficients were .52, .45, and .63.

Just what abilities the tests do measure is difficult to say. The coefficients of correlation of the composite score on the three with other composite scores, intelligence, and reading are shown on Table 18, Chapter V. Considered singly, Critical Thinking showed the highest correlation with reading and IQ, about .5 in both cases. These are comparable to the coefficients Watson and Glaser (1952) report for their critical thinking appraisal, one part of which, Inference, is very similar to the present test.

Estimation is similar to Thurstone's (1938) estimating test which is described as a test of numerical reasoning. In Thurstone's analysis, however, the test had little in common with tests of numerical judgment and arithmetic reasoning. It seems to us that success on Estimation depends on a fund of practical information and skill in weighing and synthesizing it.

---

*Both Estimation and Best Answer were rather extensively tried out, with some interesting results. The progressions of total-test means and item difficulties of Best Answer were consistent and marked in samples of superior seventh, eighth, and ninth-grade students, college freshmen, and graduate students. In one of the tryouts of Estimation, two groups of about 30 eighth-grade students each were equated. One group was given an hour's instruction and practice on how to estimate before taking the test. The difference between the means of the two groups was negligible.

Curiously enough, the coefficient of correlation of Best Answer with Reasons II (see Table 9) was only about .2. Producing good reasons is apparently independent of judging given reasons. Success on Best Answer probably depends both on information about the subject matter and evaluation of past experience.

All three tests would appear to be concerned with the ability to make judgments of complex stimulus materials, several aspects of which must be taken into account simultaneously. Johnson (1955) believes that the important abilities required for judging are (1) abstracting the pertinent data from the complex situation, (2) adopting and maintaining a set for these pertinent data, (3) attending to several data simultaneously, (4) weighing each appropriately, (5) integrating the information so that it is related to one of the response alternatives rather than others, and (6) exerting caution or deliberateness. Guilford (1957) lists practical judgment, practical estimation, and logical reasoning among his evaluation abilities.

Although it is by no means clear just what simple abilities the three tests evoke, it is clear that two of the three differentiate good and poor problem-solvers sharply. The former are considerably better at judging the accuracy, goodness, or pertinence of given information.

Differences in Preferences

The means of good and poor problem-solvers on Preferences were 22.56 and 20.93, respectively. The t value of the difference was 1.69 with 213 degrees freedom, barely significant at the 10 per cent level. The reliability of the test, estimated from half-test scores in a random sample of 118 tests, was .82.

In our tryouts, none of the coefficients of correlation of the two preliminary forms of Preferences with preliminary forms of Thought Problems, in samples of 26 or more, was significant at the 5 per cent level. The coefficient of correlation of Preferences with IQ was significant in only one sample.

Thus, there is considerable evidence that preference for thinking activities, as measured, does not differentiate good and poor problem-solvers. However, we doubt that our measure of preference is valid enough to justify acceptance of the hypothesis that good and poor problem-solvers do not differ in liking to think. Activities of the kind included in Preferences may be preferred or not preferred for various reasons indifferently related to liking or not liking to think.

It will be recalled that the memory test included in the question, "Did you talk about any of the problems with someone outside of class?" and that the difference between percentages of good and poor problem-solvers answering "yes" was highly significant. This would seem to be an indication of greater interest in or concern about thinking activities on the part of the good problem-solvers, but the evidence is far from direct.*

Differences in Variability

On five tests, Planning II, Consequences II, Concealed Figures, Preferences and Critical Thinking, the good problem-solvers were significantly more variable than the poor, at the 5 per cent or lower level; on an equal number of tests, Matrices, Seeing Problems II, Missing Facts, Mutilated Pictures, and Best Answer, they were significantly less variable. We studies these tests and the abilities which seemed to be involved to see whether there were logical explanations of the heterogeneity. None of the explanations that occured to us would explain the differences, i.e., would have enabled us to predict that the good (or poor) problem-solvers would be significantly greater in variability in this sort of test, significantly less in that.

---

*One of our colleagues, Mr. Joseph Moore, is making an extensive study of cultural differences between the good and poor problem-solvers with IQ of 120 or more in the Lansdowne sample. He reports that the two groups are significantly different on the Kuder Preference Record only in the Scientific category. (He also reports that the educational level of the parents of the good problem-solvers is significantly higher than that of the parents of the poor, and that a significantly greater number of good problem-solvers come from homes which can be classified democratic as opposed to autocratic.)

Significant differences in variability appear to have little effect on the t test in large samples, but they do raise common-sense questions about the fairness of averages in comparing groups. The matter is related to overlap of sets of data under statistical comparisons. Statistical tests establish significance or lack of significance of differences between averages of groups, but they are not directly concerned with overlap.

Overlap and differences in variability, however, are of great concern when one is trying to isolate the components of an ability, such as problem-solving. The fact that numerous poor problem-solvers are above the mean of good problem-solvers in the various comparisons reported in this chapter makes it difficult to come to any conclusions about critical abilities in, or necessary conditions of, problem solving.

We are planning to study the profiles of test scores and combinations of scores of good and poor problem-solvers to see whether patterns of scores are related to success in problem solving. The results of the study will be reported in another paper.

Sex Differences

In addition to the Mechanical Reasoning and Davis-Eells tests, previously noted, boys were significantly better than girls, at the 5 per cent or lower level, in Consequences I, Preferences, and Best Answer. Only Mechanical Reasoning and Preferences, however, showed large differences. On the former, the boys' mean was 32.2, the girls' mean 23.3; on the latter, the boys' 24.8, the girls' 19.5.

Girls were significantly better in Words, First Letter, First and Last Letter, Memory, and Planning II, but only First Letter showed a large difference. For that test, the girls' mean was 52.2 and the boys' 44.9.

These differences were not unexpected; in fact, in view of the amount of literature on sex differences in reasoning and problem solving, we had expected more decisive ones. The size and direction of the differences indicate only that boys are better than girls at solving problems requiring mechanical reasoning, while girls are better than boys on word-association problems. With respect to general reasoning or problem-solving abilities, there is little evidence in our study that either sex is better than the other.

Differences in Errors

Much of what is known or believed about problem-solving skills and processes is, of course, based on observing how people go about solving problems. The studies of Maier (1933), Chrisof (1939), Duncker (1945), Bloom and Broder (1950), Burack (1950), Marks (1951), and Buswell (1956) are illustrative of the many which analyze successful and unsuccessful attempts to solve given problems. Although not primarily concerned with errors, such studies usually attempt to classify and explain the errors in the "thinking-aloud" and other protocols for their subjects.

There have been several experimental studies directly concerned with cognitive error, illustrative of which are those of Brigham (1932) and Lawrence (1957). Brigham analyzed errors of large samples of college applicants on various kinds of items of the Scholastic Aptitude test of the College Entrance Examination Board. He found that a frequent source of error appeared to be the distortion of the specific requirements of problems to more general, less difficult requirements and that erroneous answers to problems having very large numbers of probable answers were concentrated at definite points. He concluded that errors are merely characteristic solutions of the problems set and that error is not capricious but orderly.

Lawrence quantitatively analyzed the errors of 274 children (median age, 11 years, 8 months; mean IQ, 104) on the Otis group verbal test. He then selected a sample of 84 for more intensive qualitative analysis, based on the reworking of all items which were wrongly answered. He found that 39 per cent of the errors, over all IQ groups, could be traced to inadequate formulation of the task, as revealed by the substitution of a general requirement for the more demanding specific requirement of the task and by other forms of difficulty reduction. He classified other conditions producing error as prepotency of particular details or characteristics; misreading, or failure to read all the choices offered; correct working but incorrect expression of the answer; difficulty in "recentring"; lack of confidence; correct answer but for wrong reason and rejection of the correct answer on reasonable grounds; lack of specific information; and complete failure to formulate the task.

It seems important to ask whether good and poor problem-solvers, classified as such independently of the problems in hand, differ in errors on objective test items. To date, we have examined the errors of our samples on the Differential Aptitude Tests of verbal, abstract, and mechanical reasoning; Davis-Eells; Matrices; and Critical Thinking. The results are interesting.

Twenty of the 45 Verbal Reasoning items difficult enough to afford comparisons brought out differences between good and poor poor problem-solvers, significant at the 5 per cent or lower level; 19 of the 37 Abstract Reasoning items; 12 of the 40 Matrices items; and 19 of the 83 Critical Thinking items. (Only 4 of the 34 Davis-Eells items and 6 of the 57 Mechanical Reasoning items difficult enough to afford comparisons brought out significant differences--little more than the number expected under the 5 per cent level when the null hypothesis is in fact true.)

Our preliminary analyses of the errors suggest that good problem-solvers show more caution than the poor, are less prone to generalize loosely, less likely to make "stupid" errors, less likely to be misled by "noisy" or irrelevant distractors, and less likely to reduce difficulty by working easier problems than those posed, e.g., less likely to abstract a single trend in completing rows or columns of figures displaying a double trend or to complete an analogy by reasoning "this goes with that" when the correct relation is "a function of this is like a function of that."

We plan to study the differences in errors on various other tests and on the individual tests discussed in the next chapter, and to report the results in a later paper.

CHAPTER IV

DIFFERENCES BETWEEN GOOD AND POOR PROBLEM-SOLVERS (CONTINUED)

Fifty of the 117 good problem-solvers and 50 of the 117 poor, discussed in previous chapters, were selected partly at random (balance was maintained between grades, schools, and sexes) for individual testing. All of the former and 48 of the latter completed the tests and are the numbers involved in the comparisons of this chapter.

The sample of good problem-solvers included 24 boys and 26 girls with mean IQ of 116.3. The sample of poor problem-solvers included 24 boys and 24 girls with mean IQ of 114.3. The numbers in both groups from the two grades and two schools were approximately equal.

The individual tests included three "rigidity" tests, similarities, persistence, free association, "concrete reasoning", and two tests of "scientific thinking", nine tests in all.* We do not know the reliabilities of any of the tests and suspect some have little. This is a distinct limitation, though on the conservative side. As Ray (1955) remarks, "Unreliability of tasks will presumably not lead to unwarranted generalizations, but may result in failure to find relationships which would otherwise appear."

The purpose of the tests was twofold, (1) to look for further differences between good and poor problem-solvers, and (2) to see whether the two groups differed on several concrete tasks as they had on the paper and pencil tests discussed in the preceding chapter.

Administration of Tests

Rooms where distractions would be at a minimum were provided by the principal of each school. Conditions were not ideal, but distractions were neither frequent nor serious enough to affect results materially.

---

*Miss Stanier assumed responsibility for the selection, administration and much of the interpretation of the individual tests.

Before the testing began, each subject was told that the University of Pennsylvania was conducting an experiment to see how junior high school pupils worked certain kinds of problems and that the government was helping with the cost of the experiment because it, too, was interested in knowing how people go about solving problems. Each subject was instructed,

(1) Tell us if you have ever seen or heard any of these problems before.

(2) Try to do your very best.

(3) Please do not discuss the tests with any of your friends. If you talk about these problems, it would spoil the fun of other pupils and spoil the experiment.

At the end of the period, the subjects were again reminded not to talk about the problems. So far as we could tell from the test results and from talking with teachers and principals, the subjects followed the instructions faithfully.

There were nine individual tests. It had originally been planned to have the first four tests given by one examiner, and the last five by a second. After about three-fourths of the subjects had been tested, however, one examiner was unable to continue, and all nine tests were thereafter given by the same examiner. The results did not appear to be affected by the change. At the time of testing, the examiners did not know whether subjects had been classified as good or as poor problem solvers.

About three hours were required for the tests. A brief rest period was allowed at the end of the fourth test. Almost all of the subjects were extremely cooperative and seemed to find the tests highly interesting. We believe the results reflect the best efforts of the great majority.

The tests and the differences they bring out between good and poor problem-solvers are discussed below in the order in which the tests were given.

Einstellung Tests

The work of Maier and his students has stimulated a lively interest in the importance of flexibility or absence of rigidity in problem-solving behavior. Despite many studies, the question is not yet resolved whether the flexibility-rigidity variables are general traits or merely kinds of behavior evoked by given situations. In one of the latest and most comprehensive studies, Fink (1957) found no evidence of generalized rigidity effects in two samples of junior high school boys on tests of spatial, numerical, and verbal reasoning.

It is pertinent to ask whether good and poor problem-solvers, classified as such a priori, differ in rigidity, when rigidity is defined in Fink's words as "resistence to change from an established problem-solution pattern when that pattern is no longer required."

Method. Our measures of rigidity included six of the Einstellung tests used by Fink. Each test consists of seven items similar in general appearance or directions. The first six items are supposed to establish set by requiring the same method of solution; the seventh (extinction) item requires a different method. The score on each test is the ratio (rigidity ratio) of the median time taken to solve the set items to the time on the extinction item.

The first test, Outside Form, consists of seven sets of four cards, on each of which is drawn a large geometric figure with a smaller figure inside. The subject must group each set of cards into two pairs, putting similar cards together. The basis for grouping the first six sets is the large figure; the basis for the seventh, the small figure. In the Blocks test, Kohs color blocks are used. The subject is required to reproduce exactly the two-color designs drawn on seven cards. The first six cards have a similar pattern with one color slanting from upper right to lower left; the seventh has a diagonal in the opposite direction. In Figures the subject is asked to indicate by a line how a geometric figure would have to be cut to make two given figures. In the six set items a horizontal line is required; in the seventh, a vertical line. In Number Series, the subject is asked to decide what pattern is established by a series of digits and then give the next number. The set items use patterns of addition; the extinction item requires subtraction. The Water Jar test consists of set and extinction items taken from Luchins' well-known tests. The Disarranged Words test requires the subject to write the common word that can be made up from a group of scrambled letters. The set items are solved by interchanging the first two letters; in the seventh, the last two letters must be interchanged.

Results. The significance of the differences between the rigidity ratios of good and poor problem-solvers on each of the six tests was determined by Wilcoxon's rank-sum test. Subjects who had been unable to solve extinction items within two and one-half minutes were assigned the last rank. The good problem-solvers were significantly better on Water Jar and Figures at the 1 per cent level, the z values from the rank-sum test being 2.74 and 2.60, respectively. They were significantly better on Outside Form at the 5 per cent level, with z equal to 2.16, and better on Numbers at the 10 per cent level with z equal to 1.88. The groups did not differ significantly on Blocks and Disarranged Words, the two easiest tests.

It will be noticed that the tests differentiate the two groups about in order of their difficulty or complexity. This suggests that flexibility, as evidenced in breaking sets from patterns of a low order of complexity, is not related in any significant way to the flexibility needed in problem solving. This interpretation is confirmed by the results of the Color Naming test discussed below.

Color Naming

Color Naming, devised by Thurstone, appears to be a test of the ability to make responses which conflict with well-established habits and hence provides a measure of simple rigidity or perseveration.

Method. The subject is required to perform two tasks as rapidly as possible. First, he calls off from a card the colors of 100 red, yellow, green, and blue dots; then, from a second card, the colors in which 100 color-names are printed. In other words, he is not to give the names of the colors he sees spelled out, but to state whether the word is printed in red, yellow, green, or blue. The slower the subject is on the second part, relative to the first, the greater the indication of perseveration.

Results. The mean ratios of times of good and poor problem-solvers on the two parts were not significantly different. These results are in line with those reporting no relationship between perseveration and flexibility in problem solving. As noted above, the flexibility required to change sets of low-order complexity is apparently not related to problem solving in our population.

Similarities

In discussing his intelligence scale, Wechsler (1944) notes that his Similarities test is one of the best subtests in the battery. He believes that it measures the ability to distinguish between essential and superficial likenesses. Guilford (1957) reports that a test like Similarities (the main difference being that the subject is required to give six responses instead of one) has a high loading on "penetration", the ability to see beyond the immediate and obvious. It would seem that such a test should sharply differentiate good and poor problem-solvers.

Method. We selected 12 pairs of words at different levels of difficulty from Forms I and II of the Wechsler Scale whose content seemed appropriate for seventh and eighth-grade pupils: plum-peach, coat-dress, cat-mouse, piano-violin, daily paper-radio, pound-yard, mountain-lake, egg-seed, poem-statue, salt-water, praise-punishment, fly-tree. The test was given and scored according to Wechsler's directions.

Results. Scores of the good problem-solvers ranged from 8 to 21 with a mean of 14.5; scores of the poor ranged from 4 to 18 with a mean of 11.1. The t value of the difference was 5.11 with 96 degrees of freedom, very highly significant. The two groups are markedly different in the ability to distinguish between basic and superficial likenesses. The results are consistent with some of the results of the II tests discussed in the preceding chapter and connect with our brief comments about errors.

Disc-Transfer Problem

The Disc-Transfer problem is listed by Ray (1955) as a complex task useful in problem-solving research. The problem has been known for at least thirty years and has been used by several investigators. It has several advantages. Records of moves made by the subjects can easily be kept for later analysis; the problem can be made more complex simply by adding discs and is easily understood. It appears to be a test of concrete

2

reasoning that can be made as difficult as desired without adding anything new. For these reasons we included it in the battery of individual tests.

Method. The subject was shown a piece of cardboard on which three circles had been drawn. Two small discs of different sizes were placed on one of the circles, with smallest disc on top. The following directions were then given:

> Let's look at this problem. Here are three circles, A, B, and C, and some discs of different sizes. Notice that the discs are on circle A and are arranged in order of size--the smallest one on top. You are to get the discs over to circle C so that they will be in the same order on C as they now are on A. The problem has two rules, (1) you can move only one disc at a time, and (2) never place a larger disc on a smaller disc. You may use circle B as a kind of transfer station. Any questions?

After the sample problem had been solved, three discs were placed on circle A in order of size, and the subject was told to follow the same directions to solve the problem. Upon the successful solution of each problem, the number of discs was increased by one until six discs were used. The subject was allowed 20 minutes to solve as many of the four problems as he could, and all of his moves were recorded.

Results. The data in Table 12 show that the good problem-solvers were able to complete more problems in fewer moves than the poor, although not until five discs were used did the differences between percentages or median number of moves become significant at the 5 per cent or lower level.

The subjects gave little evidence of forming any sort of general rule for solving the series. Had we mentioned the possibility of a rule in the directions, behavior might have been different. Although the good problem-solvers were more efficient than the poor, the median number of moves taken by each group was far from minimum. Sixteen of the former finished the three-disc problem in the minimum number of moves, while only two of the latter were able to do so. Two of the good and none of the poor problem-solvers worked the four-disc problem in the

Table 12

Number and Percentage of Good and Poor Problem-Solvers Successful on Disc Transfer and Median Number of Moves

Number of Discs	Good Problem-Solvers N	%	Mdn	Poor Problem-Solvers N	%	Mdn
3	50	100.0	10	47	97.9	13
4	49	98.0	24	46	95.8	26
5	46	92.0	71	42	87.5	86
6	25	50.0	116	11	22.9	150

minimum number of moves. No other optimum solutions were achieved by members of either group.

It would appear that concrete reasoning tasks, such as the disc transfer problems, must involve a relatively high degree of complexity before they differentiate good and poor problem-solvers.

Persistence Tests

Persistence is presumably important in problem solving. In his summary of the research relating to persistence, Johnson (1955) writes,

> The picture is not yet clear, but the evidence points to a persistence factor that determines the amount of time people will spend on various tasks, and thus the amount they will accomplish, when the instructions about stopping are indefinite . . . . According to the evidence, it is independent of the content of the tasks and independent of general intelligence. It seems to be different also from such related factors as perseveration, rigidity . . . (p. 96)

Several investigators have noticed that willingness to spend time on a problem may be a determining factor in problem-solving success. For example, Bloom and Broder (1950) found, among other things, that non-successful problem solvers gave little time to the consideration of the problem and were inclined to decide rather quickly that they could or could not solve a problem.

When several problems claim an individual's attention simultaneously, his behavior is no doubt a function of the importance or interest of a particular problem to him. However, if the problems are similar in content, it seems reasonable to assume that behavior will vary as a result of such variables as persistence (desire to complete an unfinished task) and concentration.

Method. In three ten-minute sessions, each subject was tested on three sets of problems. Four anagrams were given first, then three concrete problems involving the construction of squares, and finally three tracing problems. At the start of each session, the subject was instructed to switch to a different problem whenever he solved one, when he felt that he could not solve the one before him, or when he just felt like switching. Each problem could be attempted any number of times, and as many problems as possible were to be solved. A record was kept of each time a switch was made.

None of the problems was easy for our subjects, and no more than one-third solved any one. The anagrams included NMGOINR (morning), EVSUORN (nervous), BBRCLIKDA (blackbird), and MTARU (impossible to solve). These were printed on different colored cards so that the examiner could tell when switches were made.

The pieces for the construction problems were cut from different colors of heavy cardboard. They were placed in three piles in front of the subject who was told that the pieces in each pile could be put together to form a perfect square.

One of the tracing problems required the subject to connect 9 dots arranged in a square with four straight lines; a second, to connect 16 dots arranged in a square to form a given figure without retracing or lifting his pencil; the third, to connect certain of 25 dots arranged in a square in such a way as to form a cross with 5 dots inside and 8 dots outside. Each series of dots had been reproduced many times on a separate sheet.

Results. The fact that significantly greater percentages of good problem-solvers worked one or more of the anagram, construction, or tracing problems, confirms once more our classification of them. As a matter of fact, five of the poor problem-solvers accounted for more than half of the solutions by their group. However, our main interest was in the number of switches made by the two groups. The statistics are shown in Table 13. It will be noted that the good problem-solvers switched fewer times on all three tests. Only on the construction problem does the difference fall short of statistical significance at the 5 per cent level.

Table 13

Mean Number of Switches by Good and Poor Problem-Solvers on the Persistence Tests

Test	Mean GPS	 PPS	Mean Diff.	t Value of Diff.	Level of Sig.
Anagrams	10.4	13.1	2.7	2.21	.05
Construction	3.0	3.6	.6	1.86	.10
Tracing	2.4	3.0	.6	2.12	.05

The three types of problems apparently aroused different degrees of interest, since fewer switches occurred on the construction and tracing problems. However, the fact that poor problem-solvers tend to make more switches regardless of the type of the problem suggests that persistence, as measured, may differentiate the groups independently of the kind of task in hand.

It should be noted that the situations were not exactly the same for the two groups, since the good problem-solvers experienced somewhat more success in working the problems. The results might have been different if all of the problems had been impossible, or if the subjects had been told that there was no time limit and that they could give up when they wished. However, our observation of the subjects at work leads us to believe that a better controlled experiment would yield more pronounced differences.

Free Association

In their study, Wilson et al (1953) defined originality as "the ability to produce ideas that are statistically infrequent for the population of which the individual is a member." One of the tests showing a high loading on originality as defined was a group test, Quick Responses, which required the subject to write down the first word he thought of in response to each of 50 stimulus words. Guilford et al (1956a) in another factor analysis found that Quick Responses did not have a significant loading on originality. However, they believed that this might have been due to the fact that the scoring key of statistical frequencies of the responses was derived from another sample.

To determine whether our good problem-solvers would give more uncommon responses than the poor, we chose 50 words from the Kent-Rosanoff Word Association Test.

Method. Each subject was instructed to listen to a word spoken by the examiner and then to say the first word he thought of. Only immediate responses were counted in the scoring. The frequencies of all responses were tabulated, and the score for each response was the percentage of subjects with the same responses. For example, if 50 per cent of the subjects said "rug" in response to "carpet", the response "rug" got a score of 50. Average scores were then computed for each subject. The lower the average score, the more uncommon the subject's responses.

Results. The mean of the good problem-solvers was 33.0, that of the poor 35.1. While the difference is in the expected direction, it did not approach significance, with P greater than .30. It may be that originality is not related to problem solving as we have defined it, or it may be that infrequency of responses in the free-association test is not an appropriate measure.

Box Problem

According to Taylor and McNemar (1955), the various experiments which have been carried out leave no doubt as to the reality of the phenomenon which Duncker has described as functional fixedness. They believe that it is operationally similar to the inhibiting effects of set and classify it with the flexibility-rigidity variables. In Duncker's (1945) words,

> . . . For the psychology of thinking, there hardly exist more fundamental differences among the various relevant elements of a problem-situation than those which determine how easily . . . they may be recognized as solution material. (p. 85)

We hypothesized that functional fixedness or the inability to see other than conventional uses for common objects would differentiate our good and poor problem-solvers sharply. We selected Duncker's box problem both because it has been used by several investigators and because it is very simple to give.

Method. The materials for the problem were placed in certain fixed positions on a table before the subject. They included three boxes, five matches, three small candles, five thumbtacks, two pieces of wood, two corks, several pencils, four paper clips, two pieces of tinfoil, and two pieces of string. The tacks, candles, and matches were placed inside the box. The subject was instructed,

> You are going to be given a problem. In order to solve it, you will need to use certain objects which are among those on the table. You may use anything on this table in any way you like. Here is the problem. Figure out a way to set three candles upright on that door at eye level (Examiner showed height on door) so they can burn easily.

The directions were repeated as often as necessary. Subjects were allowed ten minutes to solve the problem and were given complete freedom of movement and use of materials.

Results. Solutions which fulfilled the standards set up in the directions were given by 25 good and 13 poor problem-solvers. The difference between proportions successful is significant at the 5 per cent level. The median time for solution of the good was 3'40", and for the poor 7'00". The median difference is significant at the .1 per cent level.

A total of 18 different solutions, ranging from the classic one proposed by Duncker to Rube Goldberg contrivances, were given. The boxes were used in nine different solutions of the good and six of the poor problem-solvers. The difference is not significant.

The results from this problem are difficult to interpret. The problem has too many solutions. It is impossible to tell just how much overcoming of functional fixedness or other inflexibilities occurred in either group. The only conclusion we can draw is that good problem-solvers are more resourceful and faster in handling concrete materials in solving problems like the box problem.

Experimentation

As noted earlier, p. 2, Dewey's logical analysis of problem solving or reflective thinking provides a formal description of "scientific thinking" or scientific method. We were interested to know whether our seventh and eighth-grade pupils would think scientifically about superstitions and certain problems, and whether such thinking would differentiate the good and poor problem-solvers. Accordingly, we devised some exercises which, for want of a better name, we called Experimentation.

Method. The exercises below were read separately to each subject, and his responses were recorded verbatim. The subject was urged to tell as much as he possibly could about his experiment. He was questioned only when one of his statements was not clear.

(1) Suppose you met someone who warned you that breaking a mirror meant seven years bad luck. This person would not be convinced by what science books or teachers said about the belief. What experiment could you do which would prove to him that the belief was false?

(2) There is a belief that if a black cat crosses your path, you will have bad luck. If you were a scientist, how would you go about actually finding out if the belief were true or false?

(3) It is a well-known fact that plants need light in order to live. Suppose you were teaching kindergarten and some of the boys and girls didn't understand or believe that plants need light. Can you think of an experiment that you could do which would prove that plants need light?

(4) Suppose you have some rats and are interested in learning about them. How would you go about proving whether hunger or thirst was stronger in a rat?

(5) There is a saying that red-haired people have bad tempers. Suppose you were made chairman of a committee in science class to investigate the truth of the saying. How would you go about actually finding out in an experiment whether the saying was true or false?

Method. The subjects' answers to each exercise were scored both by sorting into five piles on the basis of quality and by assigning scores from 0 to 5 on the basis of important points mentioned. The two methods gave approximately equivalent results. Answers varied from brief statements to long and sometimes discursive discussion. There appeared to be no relation between length and quality of answers.

The means of the scores obtained by the second method of scoring of the two groups on the five exercises are shown in Table 14. In an overall three-way analysis of variance, with the 48 poor problem-solvers paired with 48 of the 50 good on the basis of IQ, the differences between groups, exercises, and individuals were all significant at the .001 level. However, the reliability estimate of the exercises as a whole computed from the F ratio for individual effects was only .20.

Table 14

Mean Scores of Good and Poor Problem-Solvers on Five Experimentation Exercises and Significance of Differences

Exercise	Mean GPS	Mean PPS	Mean Diff.	t Value of Diff.	d.f.	Level of Sig.
1. Mirror	1.44	1.19	.25	1.56	96	.20
2. Black Cat	2.00	1.48	.52	2.27	96	.05
3. Plants	2.67	2.46	.21	.87	96	.30
4. Rats	2.23	1.40	.83	3.53	96	.001
5. Red Hair	2.15	1.87	.28	1.33	96	.20

Because of the low reliability of the exercises, considered as one test, and the highly significant differences between exercises, we hesitate to draw any conclusions. Although the good problem-solvers appear to be significantly better than the poor, the exercises were too hard for both groups. Fewer than half of the total sample were able to give an answer which could be considered scientifically adequate to even Exercise 3, a type of experiment usually included in elementary science courses.

Problem-Solving Apparatus

Exercises 7, 16, 31, and 35 of Thought Problems (see Appendix) demand a process of hypothesis verification and thus connect with a step in scientific thinking. We constructed an electrical apparatus which made it possible to present similar problems concretely. Our hypotheses were that (1) the problems would differentiate good and poor problem-solvers and (2) the poor problem-solvers would perform better on the apparatus than on the written problems.

Before discussing method and results, we should pause to note that such problems are essentially concerned with how information is obtained and/or used in problem solving. They are similar in some respects to those involved in the tab-item and trouble-shooting techniques of Glaser et all (1954) and Rimoldi (1955); in John's PSI apparatus; and in the concept attainment materials of Bruner et al (1955).

Method. The apparatus consists of a 30" by 20" panel with twelve sockets or switches at either end in which brass balls complete electric circuits. At one end are 12 lights, each of which is connected in series to an upper switch and a lower, but not corresponding, switch. A light comes on when balls are placed in the two appropriate switches. At the end opposite the lights, the switches are lettered A through L, marked off in groups of three. There is also a light at the left center of the board which comes on when a ball is placed in either switch B or D; a light at the middle center which comes on when balls are placed in both switches F and H; and a light at the right center which comes on when a ball is placed in either switch I or J. The circuits for the middle three lights are independent of each other and of the circuits for the twelve upper lights. A master multiple switch controls all circuits.

In Part I of the apparatus experiment, the subject was informed that the switches in front of him were connected by concealed wires to the lights at the opposite end of the board, and that any light could be turned off by removing a certain ball. He was told that his ability to follow directions was about to be tested. The instructions below were read aloud, while the subject read silently a copy of his own,

(1) Think of the lights as being numbered 1 to 12 from left to right. (Examiner illustrated left to right.)

(2) When a light comes on, decide what number it is.

(3) Find that number on this strip. (A strip of scrambled numbers from 1 to 12 was then placed directly in front of subject, the numbers lining up with switches).

(4) The correct ball to remove is the ball directly below the number selected on the strip.

These directions could be referred to at any time. Subjects had to succeed on a sample problem before being permitted to proceed. They received help on the sample if necessary.

The experiment involved three stages. During Stage I the strip of numbers was directly in front of the subject adjacent to his switches; in Stage II the strip was placed near the middle of the board; and in Stage III it was placed at the opposite end of the board adjacent to the lights.

Each time the strip was moved, the subject was told to continue to follow the same directions. Two different problems were presented at each stage. The subject had immediate knowledge of his results, since the light in question went off immediately when the proper ball was removed. The number of errors made at each stage was recorded.

Part II of the apparatus experiment consisted of three problems each of which required the subject to discover which switches were connected to a light in the center of the board. In the first problem, the subject was told to look at switches A through F and the following instructions were read by the examiner,

> Now look at these six switches. Either one of two switches will light this light (left-center light). That is, there are two switches connected to this light and either one of them will turn it on. The switches work independently.
>
> I'm going to cover the light and give you three brass balls to put in these six switches. When you have all three in place, I'll uncover the light so you can see whether it is on. Then I'll cover the light and you can place the three balls again, and so on. We'll repeat the process until you are sure, positive that you know which two switches turn the light on. Try to find out in as few trials as possible, without guessing.

The second problem involved the six switches D through I and the center light. The instructions were,

> Now look at the six center switches. To light this center light, two balls must be placed at once. Do you understand? Both balls must be in place before the light will go on. Each time you place three balls I'll let you see whether the light is on. We'll proceed as before. Try to find out in as few trials as possible, without guessing.

The third problem involved the six switches G through L and the right center light. The instructions were identical with those for the first problem.

Records of the combinations of switches tested by the subject in reaching a solution were kept. If the subject did not appear to be reaching a solution after six minutes, the problem was discontinued and a new one begun.

Results. Part I. The frequencies of errors made by the two groups at the three stages of the first experiment are shown in Table 15. It will be noted that errors increased as the strip of numbers was moved further from the subject. The combined results confirm those of Kay (1954) who had experimented similarly with unclassified subjects. Kay suggests,

Table 15

Frequency of Errors by Good and Poor Problem-Solvers at Each Stage of First Apparatus Problem and Significance of Differences

Stage	Group	Number of Errors 0	1	2 or more	Chi Square	Level of Significance
I	GPS	38	10	2		
	PPS	36	10	2	.4	.75
II	GPS	36	12	2		
	PPS	28	11	9	5.4	.10
III	GPS	36	12	2		
	PPS	24	11	13	10.5	.01

> . . . At Stage III, the difficulty was not so much one of an inability to apply an instruction because it was "forgotten", as an inability caused by one piece of information taking priority over another; that is in this case, the apparently contrary information from the spatial proximity of the lights and the index card (strip of numbers) overriding the original instructions. (p. 167)

The interesting thing is that good and poor problem-solvers differ very significantly in frequency of error at Stage III and differ at the

10 per cent level at Stage II. If Kay's explanation is correct, our good problem solvers are significantly better at keeping information straight. At any rate, they were less confused by the changing conditions of the simple problem.

Results. Part II. Both of our hypotheses were confirmed by the results of the second part of the apparatus experiment. The numbers of solutions provided by the good and poor problem-solvers are shown in Table 16.

The former are significantly better than the latter on all three problems, and the difference on the third problem is very marked. The good problem-solvers apparently profitted from the practice on the first problem, essentially similar to the third.

Table 16

Number and Percentage of Good and Poor Problem-Solvers Successful on the Problems of the Second Part of the Apparatus Experiment.

Problem	Good Problem-Solvers		Poor Problem-Solvers		% Diff.	Level of Sig.
	N	%	N	%		
1	31	62.0	17	35.4	26.6	.05
2	48	96.0	32	66.7	29.3	.01
3	37	74.0	15	31.2	42.8	.01

The two groups showed other marked differences. A significantly greater percentage of poor problem-solvers offered solutions (guesses) before they had sufficient information and made more repetitions of combinations of switches. Their median number of repetitions on the first problem was 4; on the second, 1; and on the third, 5; compared to 1, 0, and 2, respectively, for the good problem-solvers. They varied more than one ball more frequently than the good problem-solvers, and thus showed less understanding of hypothesis verification. All in all, the three problems brought out various kinds of differences between the two groups, most of which were significant.

We are planning to analyze the moves further to determine the extent to which our seventh and eighth-grade pupils make use of some of the strategies Bruner's more sophisticated thinkers used in concept attainment.

Data supporting our second hypothesis, namely, that the poor problem solvers perform better on the apparatus than on the written problems, are shown below. The numbers passing and failing the first and second apparatus problem and Items 7 and 16 of *Thought Problems* are shown in the exhibits, the numbers of good problem-solvers being in parenthesis.

		First Apparatus Problem	
		Fail	Pass
Thought Problem, No. 7	Pass	0 (9)	2 (12)
	Fail	31 (10)	15 (19)

		Second Apparatus Problem	
		Fail	Pass
Thought Problem, No. 16	Pass	0 (1)	9 (31)
	Fail	16 (1)	23 (17)

The difficulties of Problem 7 and the first apparatus problem for the poor problem-solvers are 4 per cent and 35 per cent, respectively; for the good, 42 per cent and 62 per cent. The difficulties of Problem 16 and the second apparatus problem for the poor problem-solvers are 19 per cent and 67 per cent, respectively; for the good, 64 per cent and 96 per cent. Although both groups do significantly better on the apparatus than on the thought problems, the poor problem-solvers show markedly greater relative gain.

CHAPTER V

THE IMPORTANCE OF CERTAIN INTELLECTUAL VARIABLES IN PROBLEM SOLVING*

One of the major questions with which this study is concerned is, What is the importance of certain intellectual variables in accounting for problem-solving ability? In this chapter we consider the question from another approach.

As noted in Chapter I, we had originally planned to give the second battery of tests only to the selected good and poor problem-solvers. However, because of the administrative difficulty of separating pupils, the tests were given to all. The additional data provided an unforeseen opportunity to gain further evidence about problem-solving ability.

It is always a good question whether differences between extreme groups have the same meaning when an entire sample is considered. The advantage of having scores for all subjects was twofold, (1) we could see whether the variables which had differentiated good and poor problem solvers contributed significantly to problem-solving ability in the entire sample, and (2) we could determine the relative importance of the variables. As far as we know, their importance has not been previously investigated through multiple-regression analysis of large samples.

The tests in the second battery and distributions of the pupils' scores are discussed in Chapter II. In this chapter we discuss the method of combining scores, the reasons for choosing the tests for each variable, and the results of the correlation and regression analyses.

The Predictor Variables

Of the seven test categories in the second battery, we eliminated memory and preferences. The role of memory in problem solving is being investigated by Hoffman (1959) in another study; preferences were not considered because they reflect likes and dislikes rather than intellectual ability. The five variables in this study are problem recognition, word fluency or word association, ideational fluency, closure, and judgment. To these we added IQ and reading scores, making a total of seven predictor variables in the multiple regression equation.

---

*The analyses of this chapter were made by Mr. Harootunian.

The means and standard deviations of each of the tests representing the variables are shown in Table 17. Since the distributions of several tests were markedly nonnormal, the scores on all tests were converted to normalized standard scores by the T-score technique. A subject's composite score on each variable was obtained by averaging his standard scores on the tests included under that variable.

It should be mentioned here that the analyses of this chapter proceeded in part independently of those in Chapter III. Although the refined or II tests differentiated more sharply the good and poor problem-solvers, we used the gross or I tests in every instance because we wanted to evaluate tests similar to those employed in various factor analyses of reasoning ability. Reasons for eliminating some of the other tests are given below. Perhaps it should be emphasized that we make no claim for appropriateness of the names we have given the variables. The variables are best described in terms of the operations required by the respective tests.

Problem Recognition. Of the three possibilities, *Seeing Problems* and *Missing Facts* were selected for the problem-recognition variable. *Planning I* was not included because at the time it was raising some scoring difficulties and had not been finally scored. It proved to be the most reliable of the problem-recognition tests and brought out highly significant differences between good and poor problem-solvers. Had we known this, it of course would have been included.

Word Fluency. The tests representing word fluency or word association are *First Letter* and *First and Last Letter*. The *Words* test was dropped because it is probably a measure of free association rather than a test of word fluency. *Similar Words* was not used because it was too difficult for our sample and lacked reliability.

Ideational Fluency. The test *Consequences I* was the only one of the possibilities not used. We deleted it because it was the only test in the battery on which poor problem-solvers got significantly higher scores.

Table 17

Means and Standard Deviations of the Problem-Solving Criterion and the Tests for Seven Predictor Variables for 513 Subjects*

Variable	Mean	Standard Deviation
Problem Solving	82.50	17.06
Problem Recognition		
Seeing Problems I	20.80	8.93
Missing Facts	18.72	5.37
Word Fluency		
First Letter	48.80	13.53
First and Last Letter	13.78	5.45
Ideational Fluency		
Reasons I	20.80	8.66
Ideas I	30.04	10.60
Groups of Things	25.63	8.31
Uses I	18.82	8.60
Closure		
Incomplete Pictures	16.58	3.73
Incomplete Words	18.38	5.78
Concealed Figures	45.13	9.59
Judgment		
Estimation	27.80	5.59
Best Answer	50.02	10.51
Critical Thinking	33.97	7.32
Intelligence Quotient	113.35	12.32
Reading Grade Equivalent	8.87	1.35

*In order to keep from drastically reducing the size of the sample, subjects who had no scores on three or fewer tests were assigned the mean scores on the missed tests. The actual frequencies on the tests with less than 513 subjects are: Seeing Problems I, 501; Missing Facts, 507; First Letter, 504; First and Last Letter, 505; Reasons I, 511; Ideas I, 512; Groups of Things, 503; Uses I, 506; Incomplete Words, 512; Concealed Figures, 495; Estimation, 498; Best Answer, 482; and Critical Thinking, 464.

Whatever this test measured appeared to be negatively related to problem-solving ability. Hence, the tests of ideational fluency are Reasons I, Ideas I, Groups of Things, and Uses I.

Closure. The closure tests include Incomplete Pictures, Incomplete Words and Concealed Figures. Mutilated Pictures and Hidden Pictures were dropped from the composite. In Chapter III we expressed doubt whether the former, ac administered in this study, is a test of closure; the latter was deleted because its time allowance was insufficient and it showed poor reliability.

Judgment. All three of the judgment tests, Estimation, Best Answer, and Critical Thinking, were included in the composite. We seriously considered eliminating Estimation because of its relatively low reliability, but retained it for two reasons. First, as reported in Chapter III, the coefficients of correlation between this test and Best Answer and Critical Thinking indicate that Estimation apparently measures an aspect of judgment that has little in common with the others. Second, although the reliability of Estimation is comparatively low, its inclusion probably would add to the realiability of the composite.

Results

Product-moment coefficients of correlation between all pairs of variables were found from the normalized standard scores and are presented in Table 18. The coefficients, ranged from a low of .1688 between ideational

Table 18

Coefficients of Correlation Between the Variables

Variable	2	3	4	5	6	7	8
1. Problem Solving	.6241	.4172	.2898	.3961	.7070	.6822	.7323
2. Problem Recognition		.4008	.5629	.2940	.5518	.5194	.5981
3. Word Fluency			.4602	.3320	.3654	.4033	.4899
4. Ideational Fluency				.1688	.2812	.3178	.3627
5. Closure					.2715	.3615	.3501
6. Judgment						.5406	.6366
7. IQ							.5575
8. Reading							

fluency and closure to a high of .7323 between problem solving and reading; all are significant at the .0001 or lower level. It will be noticed that the coefficients between problem solving and the selected variables in every case are in line with the results previously brought out by the differences between good and poor problem-solvers.

As far as this chapter is concerned, the data providing the most important information are the beta coefficients, since the direct contribution of each predictor variable in a multiple regression equation equals the square of its beta coefficient. These coefficients were computed from Fisher's auxiliary statistics; through this technique one or more of the predictors can be eliminated and a new set of betas readily determined for the remaining variables. The beta coefficients, as well as the multiple coefficients, for the combinations of seven, six, and five predictor variables are shown in Table 19. We discuss the meaning of these coefficients under each of the predictor variables.

Table 19

Beta Coefficients and Multiple Coefficients of Correlation Between Problem Solving and Selected Variables

Variable	Beta Coefficient Seven Predictors	 Six Predictors	 Five Predictors
Problem Recognition	.1830	.2590	.3359
Word Fluency	.0070*	.0661*	.1100
Ideational Fluency	-.1033	-.1094	-.1087
Closure	.0795	.1013	.1520
Judgment	.2690	.3712	.4707
Intelligence Quotient	.2733	.3194	**
Reading	.3060	**	**
(Multiple Coefficient)	.8482	.8234	.7859

*Not significant at the .01 level. All other coefficients are significant at the .01 level or lower.

**Eliminated from regression equation.

Problem Recognition. The data indicate that, while problem recognition is not the most important variable, it contributes substantially to problem solving, as defined. This confirms the results of Chapter III. When IQ and reading are eliminated from the regression equation, problem recognition is the second best of the variables, accounting directly for about 11 per cent of the problem-solving variance. Had *Planning I* or *II* been included, the contribution of problem recognition would no doubt have been greater.

Word Fluency. When combined with the other variables, word fluency becomes relatively unimportant in accounting for problem solving. As a matter of fact, its *beta* coefficient is significant at the .01 level only when IQ and reading are dropped as predictors, and even then it accounts for only 1 per cent of the variance. The results for word fluency are disappointing when it is recalled that both of the tests in the composite differentiated good and poor problem-solvers significantly. Apparently, the ability to produce words rapidly according to certain requirements has a role in problem solving; but it is a minor one.

Ideational Fluency. It will be noted that the *betas* in Table 19 for ideational fluency are all negative. This means that ideational fluency, in the combination, actually interferes with prediction of problem-solving ability. We can only speculate why this is so. It may be that individuals who call up ideas in quantity at the expense of quality tend to be unable to separate the superficial from the important and, hence, tend to be poor judges of information and possible problem solutions. The fact that *Consequences I* of the second battery differentiated significantly in favor of the poor problem-solvers is consistent with this hypothesis, as are the results from the II tests of the second battery and the *Similarities* test of the individual battery. We mentioned earlier, p. 24, that it is logical to suppose that the person capable of rapidly producing ideas, other things equal, will solve more problems than one less capable. It may be that, experimentally speaking, other things are not ordinarily equal.

Closure. Although closure contributes significantly as a member of each combination, the percentage of problem-solving variance which can be attributed to it is very small, ranging from less than 1 per cent to slightly more than 2 per cent. It would seem that closure has little importance in problem-solving ability, as defined.

Judgment. Of the five composites, judgment contributes the most to the variance of problem solving. In the seven-variable combination it accounts for almost as much as IQ and reading; in the others it explains the most. In fact, when the five composites alone are considered, more than one-third of the explained variance of problem solving can be attributed directly to judgment. The importance of this variable in the multiple regression equations verifies the sharp differences found between good and poor problem-solvers on the separate tests of the judgment composite.

Intelligence. The importance of intelligence in explaining problem-solving ability is evidenced by its two relatively high betas. These coefficients were not unexpected. In Chapter I we reported finding few good problem-solvers with IQ below 105, which necessitated the statistical control of intelligence.

Of particular interest, however, is the fact that in neither regression equation does IQ account for the greatest proportion of problem-solving variance. It is noteworthy that the elimination of reading as a predictor left the beta coefficient for IQ relatively unchanged, while others increased considerably, particularly the betas for judgment and problem recognition. As a matter of fact, judgment seems to be just as important as IQ in explaining problem solving, if not more so.

Reading. In the seven-variable combination, the greatest proportion of problem-solving variance is accounted for by reading. Earlier in this study we cited Thorndike's (1917) conclusion that reading and problem solving are essentially similar processes. The evidence from this analysis adds support to Thorndike's conclusion. It is evident that the abilities elicited by reading are fully as important in problem solving as any of the variables we have studied.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

This study has been principally concerned with the questions (1) what variables differentiate good and poor problem-solvers among seventh and eighth-grade pupils alike in respect to average IQ and (2) what is the importance of the variables in explaining or accounting for problem-solving ability.

No attempt was made to define the term problem-solving precisely or to distinguish it from the terms thinking and reasoning. The criterion measure of problem-solving ability was a composite score on four of the reasoning tests described below.

In October, 1958, the Differential Aptitude Tests of verbal, abstract, and mechanical reasoning; the Davis-Eells Games; the Standard Progressive Matrices; and a test Thought Problems consisting of 40 reasoning problems were given to 316 seventh-grade pupils in 11 sections and 320 eighth-grade pupils in 12 sections in the Abington and Lansdowne (Pennsylvania) Junior High Schools. The tests were administered by regular teachers of the schools.

The median chronological age of the 636 pupils was 154 months with decile deviation from 143 to 165 months; the median intelligence quotient on the California Test of Mental Maturity, Long Form of the 615 available from school records was 114 with decile deviation from 97 to 128. The sample appeared to be well representative of large-city suburban seventh and eighth-grade pupils.

To obtain a criterion or composite measure of problem-solving ability, the average of the scores on the Differential Aptitude Tests of verbal and abstract reasoning was added to the sum of the scores on the Davis-Eells and Thought Problems tests. The coefficient of reliability of the composite or criterion was .93.

The mean of the criterion measure for the 554 pupils having scores in all tests and recorded IQ was 82.3, and the standard deviation 16.9.

Four lines of regression of the composite on IQ were determined--one for each grade in each school--and the deviations of the criterion scores from the lines were computed. Good problem-solvers were defined as those having the largest positive deviations; poor problem-solvers as those having the largest negative deviations. From these deviates, 117 good and 117 poor problem-solvers were selected, all of whom had deviations in excess of $.65(SE)_{est}$.

The sample of 117 good problem-solvers included 27 seventh-grade boys, 25 seventh-grade girls, 31 eighth-grade boys, and 34 eighth-grade girls. The mean of the group on the criterion measure of problem-solving ability was 98.7; the average of deviations from the line of regression, +14.9; and the mean IQ, 114.3.

The sample of 117 poor problem-solvers included 30 seventh-grade boys, 29 seventh-grade girls, 33 eighth-grade boys, and 25 eighth-grade girls, The mean of the group on the criterion was 66.6; the average of deviations from the line of regression, -15.1; and the mean IQ, 113.5.

The differences between means of the good and poor problem-solvers on the Mechanical Reasoning and Matrices tests and on arithmetic and reading achievement tests were large and highly significant. In addition, the two groups differed significantly (P less than .001) in the ranks which had been assigned earlier by the mathematics teachers on the basis of resourcefulness and ability in working thought problems. It was concluded that the method of selection had resulted in 117 pupils who, as a group, were good at solving intellectual problems and in 117 pupils who were relatively poor, with intelligence statistically controlled.

It had originally been planned to give the tests listed in Table 20 to only the two groups; however, because of the administrative difficulty of segregating pupils, the tests were given to all pupils. The test statistics for the good and poor problem-solvers present on days the tests were given are shown in the table. In all cases 105 or more of each group were present. All of the tests were given during November, 1958, except Best Answer and Critical Thinking. These were given in January, 1959.

In brief description of the tests; Seeing Problems asked the subject to list problems which might come about in connection with a candle and a

Table 20

Summary of Group Test Statistics for Good and Poor Problem-Solvers*

Test	Reliability	Mean		Mean Diff.	Level of Sig.
		GPS	PPS		
Problem Recognition					
Seeing Problems I	.85	19.55	19.69	-.14	.80
Seeing Problems II	.80	12.91	10.46	2.45	.001
Missing Facts	.86	21.76	15.65	6.11	.001
Planning I	.90	21.64	18.38	3.26	.001
Planning II	.64	11.76	9.01	2.75	.001
Word Association					
Words	.73	51.80	50.50	1.30	.30
First Letter	.86	50.31	46.46	3.85	.05
First & Last Letters	.68	14.39	12.62	1.77	.01
Similar Words	.28	8.22	6.90	1.32	.001
Ideational Fluency					
Reasons I	.81	19.61	20.04	-.43	.60
Reasons II	.71	14.24	13.84	.40	.50
Ideas I	.85	29.84	29.15	.69	.50
Ideas II	.83	22.33	18.89	3.44	.001
Groups of Things	.66	25.75	23.88	1.87	.10
Consequences I	.73	20.79	26.19	-5.40	.001
Consequences II	.74	9.15	5.23	3.92	.001
Uses I	.76	17.61	19.03	-1.42	.20
Uses II	.68	13.23	11.18	2.05	.01
Closure					
Mutilated Pictures	.86	31.94	29.67	2.27	.001
Hidden Pictures	.40	6.96	6.42	.54	.20
Incomplete Pictures	.71	17.39	15.97	1.42	.01
Incomplete Words	.88	19.41	17.96	1.45	.10
Concealed Figures	.88	46.46	44.12	2.34	.10
Judgment					
Estimation	.37	28.84	27.10	1.74	.05
Best Answer	.77	55.31	45.04	10.27	.001
Critical Thinking	.66	38.17	29.92	8.25	.001
Memory	**	5.19	3.22	1.97	.001
Preferences	.82	22.56	20.93	1.63	.10

*There were at least 105 of each group in all comparisons.

**Not estimated.

pencil; *Missing Facts*, to list the facts needed to solve given arithmetic problems; and *Planning*, to state the information needed to reach decisions about given practical problems. *Words* asked the subject to list any words; *First Letter*, words beginning with given letters; *First and Last Letters*, words beginning and ending with given letters; and *Similar Words*, synonyms of the words "large" and "fast". *Reasons* asked the subject to give reasons why people like to have trees in their yards and like to live in cities; *Ideas*, to list ideas related to given topics; *Groups of Things*, to list things that are round and things that are rectangular; *Consequences*, to state the consequences of two imagined events; and *Uses*, to list possible uses of a newspaper and a brick. *Mutilated Pictures* asked the subject to mark the picture in given pairs of pictures in which a part is missing; *Hidden Pictures*, to find faces embedded in pictured scenes; *Incomplete Pictures*, to identify objects from incomplete pictures; *Incomplete Words*, to identify words from given parts of letters; *Concealed Figures*, to identify simple designs embedded in complex designs. *Estimation* asked the subject to make quantitative estimates in given situations; *Best Answer*, to select the best and next-best answers to given practical questions; and *Critical Thinking*, to indicate whether inferences from given information were logical. *Memory* asked the subject to list as many problems as he remembered from the *Thought Problems* test taken two weeks earlier; and *Preferences*, to select the preferred activity in pairs of activities, one of which would ordinarily involve more thinking than the other. (It was doubtful whether the last two tests, in particular, measured what they were designed to measure.)

The scores on the I tests listed in Table 20, except *Consequences*, were the number of different, sensible items listed; the scores on the corresponding II tests, the number of independent, clearly relevant or useful items. The II scores thus reflect quality of response. The score on *Consequences I* was the number of immediate consequences listed; the score on *Consequences II*, the number of remote.

Fifty good and 50 poor problem solvers were selected partly at random (balance between grades, schools, and sexes was maintained) for further testing by the individual tests shown in Table 21. Fifty of the former and 48 of the latter completed all tests and are the numbers involved in

Table 21

Summary of Results of Individual Tests of
50 Good and 48 Poor Problem-Solvers

Test	Results
Einstellung tests of breaking from experimentally established set.	Good problem-solvers were faster on all six tests and significantly* so on the three most complex.
Color naming test of perseveration.	No significant differences.
Similarities test of discrimination between essential and superficial likenesses.	Good problem-solvers were significantly superior. (This test was among the most discriminating of group and individual tests.)
Disc-transfer test of concrete reasoning.	Good problem-solvers solved more problems in fewer moves. As the complexity of task increased, differences became significant.
Persistence test of concentration or work habits.	In two of the three problems, good problem-solvers made significantly fewer switches from task to task.
Free Word-Association	Good problem-solvers tended to give more unusual responses, but difference was not significant.
Duncker's box problem test of functional fixedness.	Good problem-solvers were significantly more resourceful and faster, but no clear evidence of overcoming functional fixedness.
Experimentation test of scientific thinking.	Responses of good problem-solvers were, on the whole, significantly closer to the requirements of controlled experimentation.
Problem-solving apparatus test of following directions.	Good problem-solvers were significantly less confused by conflicting information.
Problem-solving apparatus test of hypothesis verification.	Good problem-solvers showed significantly more efficiency in testing hypotheses.

*At the 5 per cent or lower level. Similar meaning is implied throughout summary of results.

the comparisons. The tests were given in December, January, and February. At the time of testing, the examiners did not know whether the examinees had been classified as good or poor problem-solvers. The nature of the tests and the results are briefly summarized in the table.

As was noted, because of the administrative difficulty of segregating the good and poor problem-solvers from other pupils for testing, the tests listed in Table 20 were given to all. This provided an unforseen opportunity to determine how much the variables which differentiated good and poor problem-solvers contributed to problem-solving ability in the entire sample. Pupils who were absent for more than three tests were eliminated; pupils who had missed not more than three tests were assigned average scores for the tests missed. This had little effect on the test statistics, while making possible a sample of 513.

The correlation statistics for the 513 are brought together in Table 22, and the composition of the variables involved is summarized at the foot of the table. All distributions entering into the composites were first normalized by the T-score technique. This was desirable because of the varying shapes of the distributions. No claim is made that the variables were optimumally compounded or are properly named. What they mean can best be inferred from the operations their component tests call for.

Conclusions

It would be repetitious to itemize the various conclusions supported by the statistics summarized in Tables 20, 21, and 22. Suffice it to say that the good problem-solvers are significantly better than the poor in nearly all tests where quality of response, accuracy, or judgment is required and that, without exception, the more complex the task or the more restricted the requirements, the greater their superiority.

Taken as a whole, the results of the group and individual tests summarized in Tables 20 and 21 strongly support the hypothesis of a general reasoning or problem-solving ability operative over various kinds of laboratory or controlled-experimentation problems. It is reasonable to believe that the ability exists in other than the particular age and grade population under investigation.

Table 22

Summary of Correlation Statistics of Criterion Measure of Problem-Solving Ability and Selected Predictor Variables for 513 Subjects.

	Correlation Coefficient							Beta Coefficient		
Variable	2	3	4	5	6	7	8			
1. Problem Solving	.68	.73	.62	.42	.29	.40	.71			
2. IQ		.56	.52	.40	.32	.36	.54	.27	.32	**
3. Reading			.60	.49	.36	.35	.64	.31	**	**
4. Problem Recognition[a]				.40	.56	.29	.55	.18	.26	.34
5. Word Fluency[b]					.46	.33	.37	.01*	.07*	.11
6. Ideational Fluency[c]						.17	.28	-.10	-.11	-.11
7. Closure[d]							.27	.08	.10	.15
8. Judgment[e]								.27	.37	.47

*Not significant at the .01 level. All other coefficients are significant at the .01 or lower level.

**Variable omitted from combination.

[a] Composite of Seeing Problems I and Missing Facts.

[b] Composite of First Letter and First and Last Letters.

[c] Composite of Reasons I, Ideas I, Groups of Things, and Uses.

[d] Composite of Incomplete Pictures, Incomplete Words and Concealed Figures.

[e] Composite of Estimation, Best Answer, and Critical Thinking.

The statistics summarized in Table 22 indicate that two important components or correlatives of this general ability are problem recognition and judgment. These are themselves complex abilities, but their contributions to problem solving are independent of intelligence, as tested, and reading. Although the good and poor problem-solvers differ significantly in closure and word-fluency variables, the two in combination with reading, IQ, judgment, and problem recognition make little independent contribution to the variance of problem solving.

Another general conclusion stems from the results of the ideational fluency tests. When these tests are scored on quantity of response (as they usually are in factor analyses) they tend to discriminate in favor of the poor problem-solvers. Ideational fluency and problem solving, as defined, appear to be inversely related, if related at all. The fact that all but one of the tests, when scored on quality or relevance of response, discriminate in favor of good problem-solvers suggests that "quality-in-fluency" rather than "fluency" is the variable of concern in problem solving. (This may also be the case in creative thinking, since the latter would seem to require good as well as numerous ideas.)

The marked and highly significant differences yielded by the separate judgment tests, the contribution of the judgment variable to problem-solving variance, and the results from several of the individual tests support the hypothesis that the major difference between good and poor problem solvers is to be found in the judging or evaluating abilities and/or attitudes. The differences brought out by the two methods of scoring the ideational fluency tests appear to add evidence for the hypothesis. That the good problem-solvers do not give as many problems, reasons, consequences, and uses as the poor would seem to be due to greater concern about the quality or appropriateness of response, i.e., they are slowed down by the retroflex activity of judging. Certainly, they give better-quality and more relevant responses.

Three findings not apparent in the summary tables should be noted. First, although the boys and girls differed significantly on several tests, there was little evidence that either sex is better than the other in general reasoning or problem-solving ability. Second, analyses of errors indicated that good problem-solvers are less prone to generalize loosely and less likely to reduce difficulty by working easier problems than those posed, e.g., less likely to abstract a single trend in completing rows or columns of figures displaying a double trend or to complete an analogy by reasoning "this goes with that" when the correct relation is "a function of this is like a function of that." Third, the flexibility-rigidity variables, as measured, do not appear to differentiate good and poor problem-solvers sharply. It seems likely that these variables have achieved their reputation in problem solving where the problems were unusual or highly complex.

Concluding Remarks

Whether the general ability to solve various kinds of laboratory problems extends to daily-life problems is an important question. There are several differences between the two kinds. Daily-life problems tend to be less structured and more likely to involve emotions and different motivation variables. Moreover, it is usually necessary to solve laboratory problems in a short time, whereas many daily-life problems may be worked on and returned to at will. The probability of a satisfactory solution to a problem is no doubt partly a function of time--the time required by the "habit of delayed response" and the deliberate comparison of possible solutions. Harlow (1951) remarks,

> . . . It is believed that the habit of delaying acceptance of problem solutions is one of the important factors determining efficiency of thinking. In the human being it should be noted that . . . elicitation and selection processes may continue for days, weeks, or years. This affords a vast opportunity for previously learned habits and concepts to transfer, to generalize, and to be associated in new ways . . . (p. 499)

It would seem that the judging abilities which so sharply differentiate the good and poor problem-solvers in the present study would come into even greater play during the extended time interval of daily-life problems and that the good problem-solvers would maintain if not increase their superiority.

Intelligence is frequently defined in terms of adjustment to environment or problematic situations. Hearnshaw as quoted by Lawrence (1957) tentatively defines it as "a cluster of high grade skills concerned with problem solving." Stern's definition stressed "general mental adaptability to new problems and the conditions of life," and Binet's included "critical analysis," which would seem to be essentially similar to judging, discussed earlier. According to such definitions, the good problem-solvers are, as a group, markedly more intelligent than the poor, although their average IQ is only slightly higher. The coefficient of correlation between IQ and the measure of problem-solving ability, corrected for attenuation, is not more than .75. Hence, only a little more than half of the problem-solving variance can be theoretically accounted for by intelligence, as tested. It would seem to follow that better tests of intelligence are needed.

If intelligence is something like a cluster of high-grade skills concerned with problem solving, the possibility of its improvement seems attractive. The skills which differentiate good and poor problem-solvers in the present study appear to be the kind that can be deliberately cultivated in the classroom.

REFERENCES

Adamson, R. E. 1952. Functional fixedness as related to problem solving: a repetition of three experiments. J. Exp. Psychol. 44:288-291.

Adkins, D. C. and Lyerly, S. B. 1952. Factor Analysis of Reasoning Tests. Chapel Hill: University of North Carolina Press.

Bartlett, F. 1958. Thinking. New York: Basic Books.

Bechtoldt, H. P. 1947. Factorial investigation of the perceptual-speed factor. Amer. Psychologist, 2:304-305.

Bloom, B. S. and Broder, L. J. 1950. Problem-Solving Processes of College Students. Chicago: University of Chicago Press.

Botzum, W. A. 1951. A factorial study of the reasoning and closure factors. Psychometrika, 16:361-386.

Brigham, C. C. 1932. A Study of Error. Princeton, N. J.: Educational Testing Service.

Brownell, W. A. 1942. Problem solving. The Psychology of Learning, pp. 415-443. Forty-first Yearbook, NSSE. Chicago: University of Chicago Press.

Bruner, J. S., Goodnow, J. J. and Austin, G. A. 1956. A Study of Thinking. New York: Wiley.

Burack, B. 1950. The nature and efficacy of methods of attack on reasoning problems. Psychol. Monogr., Vol. 64, No. 313.

Burt, C. 1919. The development of reasoning in school children. J. Exp. Ped. and Trn. Col. Rec., 5:68-77.

Buswell, G. T. 1956. Patterns of Thinking in Problem Solving. Berkeley: University of California Press.

Cardall, A. J. 1950. Test of Practical Judgment. Chicago: Science Research Associates.

Carpenter, F. 1956. Educational significance of studies on the relation between rigidity and problem solving. Sci. Educ., 40:286-311.

Carroll, J. B. 1941. A factor analysis of verbal abilities. Psychometrika, 6:279-307.

Chrisof, C. 1939. The formulation and elaboration of thought-problems. *Amer. J. Psychol.*, 52:161-185.

Corman, B. R. 1958. Learing II: Problem solving and related topics. *Rev. Educ. Res.*, 28:459-67.

Davis, F. B. 1947. *The A.A.F. Qualifying Examinations.* Washington: Government Printing Office.

Dewey, J. 1910. *How We Think.* Boston: Heath.

Duncker, K. 1945. On problem solving. *Psychol. Monogr.*, Vol. 58, No. 270.

Fink, D. R. 1957. *A Study of the Nature of Behavioral Rigidity.* Unpublished Doctor's Study, University of Pennsylvania, Philadelphia.

Fink, D. R. 1958. Negative evidence concerning the generality of rigidity. *J. Abn. and Social Psychol.*, 57:252-254.

Glaser, R., Damrin, D. E., and Gardner, F. M. 1954. The tab item: a technique for the measurement of proficiency in diagnostic problem solving tasks. *Educ. Psychol. Measmt.*, 14:283-293.

Guetzkow, H. 1951. An analysis of the operation of set in problem-solving behavior. *J. Gen. Psychol.*, 45:219-244.

Guilford, J. P., Wilson, R. C., and Christensen, P. R. 1952. A factor-analytic study of creative thinking. *Rep. Psychol. Lab.*, No. 8, Los Angeles: University of Southern California.

Guilford, J. P. 1954. Factors in problem solving. *ARTC Instructor Journal*, 4:197-204.

Guilford, J. P., Kettner, N. W., and Christensen, P. R. 1955. A factor-analytic investigation of the factor called general reasoning. *Rep. Psychol. Lab.*, No. 14. Los Angeles: University of Southern California.

Guilford, J. P., Kettner, N. W., and Christensen, P. R. 1956a. A factor-analytic study across the domains of reasoning, creativity, and evaluation. *Rep. Psychol. Lab.*, No. 16. Los Angeles: University of Southern California.

Guilford, J. P., Kettner, N. W., and Christensen, P. R. 1956b. The nature of the general reasoning factor. *Psychol. Rev.*, 63:169-172.

Guilford, J. P., 1957. A revised study of intellect. *Rep. Psychol. Lab.*, No. 19. Los Angeles: University of Southern California.

Gullikesen, H. 1950. *Theory of Mental Tests.* New York: Wiley.

Harlow, H. F. 1951. Thinking. *Theoretical Foundations of Psychology*, pp. 452-505. New York: Van Nostrand.

Hoffman, C. B. 1959. *The Relationship of Immediate Recall, Delayed Recall, and Incidental Memory to Problem-Solving Ability*. Doctor's Study, University of Pennsylvania, Philadelphia. (In progress.)

Holzinger, K. J. and Swineford, F. 1939. A study in factor analysis: the stability of a bi-factor solution. *Sup. Educ. Monogr.*, No. 48. Chicago: University of Chicago.

Horst, P. 1936. Obtaining a composite measure from a number of different measures of the same attribute. *Psychometrika*, 1:53-60.

Humphrey, G. 1951. *Thinking*. London: Methuen.

John, E. R. 1957. Contributions to the study of the problem-solving process. *Psychol. Monogr.* Vol. 71, No. 447.

Johnson, D. M. 1944. A modern account of problem solving. *Psychol. Bull.* 41:201-229.

Johnson, D. M. 1955. *The Psychology of Thought and Judgment*. New York: Harper.

Kay, H. 1954. The effects of position in a display upon problem solving. *Quart. J. Exp. Psychol.*, 6:155-169.

Kelley, T. L. 1927. *Interpretation of Educational Measurements*. Yonkers-on Hudson, New York: World Book Co.

Lawrence, P. J. 1957. A study of cognitive error. *Brit. J. Educ. Psychol.*, 27:176-189.

Levine, A. S. 1955. Perseveration, rigidity, and persistence. *Psychol. Rep., Monogr. Suppl.*, 1:107-125.

Luchins, A. S. 1942. Mechanization in problem solving. *Psychol. Monogr.*, Vol. 24, No. 248.

Maier, N. R. F. 1933. An aspect of human reasoning. *Brit. J. Psychol.*, 24:144-155.

Marks, M. R. 1951. Problem solving as a function of the situation. *J. Exp. Psychol.* 41:74-80.

Matin, L. and Adkins, D. C. 1954. A second-order factor analysis of reasoning abilities. *Psychometrika*, 19:71-78.

McNemar, O. W. 1955. An attempt to differentiate between individuals with high and low reasoning ability. *Amer. J. Psychol.*, 68:20-36.

Pemberton, C. 1952. The closure factors related to other cognitive processes. *Psychometrika*, 17:267-288.

Ray, W. S. 1955. Complex tasks for use in human problem-solving research. *Psychol. Bull.*, 52:134-149.

Rimoldi, H. J. A. 1948. Study of some factors related to intelligence. *Psychometrika*, 13:27-47.

Rimoldi, H. J. A. 1955. A technique for the study of problem solving. *Educ. Psychol. Measmt.*, 15:450-461.

Saunders, D. R. 1955. *Personality Research Inventory*. Princeton, N. J.: Educational Testing Service.

Schroder, H. M. and Rotter, J. B. 1952. Rigidity as learned behavior. *J. Exp. Psychol.*, 44:141-150.

Stellar, E. 1957. Physiological psychology. *Ann. Rev. Psychol.*, pp. 415-436. Palo Alto, Cal.: Annual Reviews.

Tate, M. W. and Voss, C. E. 1956. A study of the Davis-Eells test of intelligence. *Harvard Educ. Rev.*, 26:374-387.

Taylor, C. W. 1947. A factorial study of fluency in writing. *Psychometrika*, 12:239-262.

Taylor, D. W. and McNemar, O. W. 1955. Problem solving and thinking. *Ann. Rev. Psychol.*, pp. 455-482. Palo Alto, Cal.: Annual Reviews.

Thorndike, R. L. 1950. How children learn the principles and techniques of problem solving. *Learning and Instruction*, pp. 192-216. Forty-ninth Yearbook, NSSE. Chicago: University of Chicago Press.

Thorndike, E. L. 1917. Reading as reasoning: a study of mistakes in paragraph reading. *J. Educ. Psychol.*, 8:323-332.

Thurstone, L. L. 1938. *Primary Mental Abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. 1944. *A Factorial Study of Perception*. Chicago: University of Chicago Press.

Thurstone, L. L. 1949. *Mechanical Aptitude III: Analysis of Group Tests*. Chicago: Psychometric Laboratory, University of Chicago.

Travers, R. M. W., Marron, J. E., and Post, A. J. 1955. *Some Conditions Affecting Quality, Consistency, and Predictability of Performance in Solving Complex Problems*. San Antonio, Texas: Lackland Air Force Base.

Watson, G. and Glaser, E. M. 1952. *Critical Thinking Appraisal Manual*. Yonkers-on-Hudson, New York: World Book Co.

Wechsler, D. 1944. *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins.

Wilson, R. C., Guilford, J. P., and Christensen, P. R. 1953. The measurement of individual differences in originality. *Psychol. Bull.*, 50:362-370.

Woodworth, R. S. 1938. *Experimental Psychology*. New York: Holt.

Woodworth, R. S. and Schlosberg, H. 1954. *Experimental Psychology*. New York: Holt.

APPENDIX

MEMORANDA AND SPECIMEN TESTS

Memorandum to Teachers, October 17, 1958

Memorandum to Teachers, November 7, 1958

Specimen Tests*

- Thought Problems I
- Thought Problems II
- Words
- First Letter
- First and Last Letters
- Similar Words
- Memory
- Reasons
- Ideas
- Seeing Problems
- Missing Facts
- Planning
- Groups of Things
- Consequences
- Uses
- Mutilated Pictures
- Hidden Pictures
- Incomplete Pictures
- Incomplete Words
- Concealed Figures
- Estimation Test
- Best Answer Test
- Critical Thinking Test
- Preferences

---

*We regret that our supply of specimen tests was exhausted before all reports were assembled. If specimen tests are not included in this volume, they may be obtained on loan from Mr. Merle W. Tate, School of Education, University of Pennsylvania, Philadelphia 4, Pennsylvania.

Philadelphia, Pennsylvania
October 17, 1958

TO: Teachers who will administer tests for study of problem-solving ability

FROM: Ethel Maw, Barbara Stanier, M. W. Tate

The tests you are asked to administer are part of a study of problem-solving ability sponsored jointly by the United States Office of Education and the University of Pennsylvania. We are having classroom teachers, rather than outsiders, administer the tests for three reasons:

1. The rapport will be better
2. The situation will be more natural
3. The pupils are likely to try harder.

The last is extremely important, since some of the tests will be rather difficult, particularly for seventh graders. As your principal will tell you, we plan to pay you for giving the tests, but the amount we can pay is small compared to the contribution your interest and help will make. We very greatly appreciate your cooperation.

The plan of the study is simple. The Differential Aptitude Tests, Davis-Eells Games, Standard Progressive Matrices, and Thought Problems will be given to some 700 seventh- and eighth-grade pupils. About two weeks later, another battery of some fifteen tests (most of them very short) will be given. Just what these tests will cover depends partly on what we find out from the first four tests. Them some 10 per cent of the pupils will be selected partly at random for individual testing. We have tried out many of the tests we are going to use and we can assure you that most of your pupils will find most of the tests interesting.

Our preliminary studies and other research lead us to believe that pupils of similar tested IQ's may differ rather widely in reasoning, practical judgment and general problem-solving abilities. One of the purposes of our study is to determine the extent to which intelligence, as tested, correlates with "practical" intelligence. Not all of the tests we will use are "practical," but there is reason to believe that all of them may indicate important factors underlying problem-solving ability. If these factors can be identified, the information will of course be useful to teachers in their job of improving problem-solving ability and useful in guidance of pupils.

Since the results from different sections and different schools are to be combined, it is essential that each teacher follow directions exactly in giving the tests. Where tests are timed, the timing should be as accurate as possible with ordinary watch and second hand. Care must be taken to see that pupils do not get help from each other. The supplementary directions on the attached sheet are intended to help in making the administration of the tests uniform.

SUPPLEMENTARY DIRECTIONS FOR FIRST FOUR TESTS

Before beginning any testing, please read these paragraphs to pupils. Change or improvise if you like, but emphasize the central ideas of the paragraphs.

> We read much these days about science and research. Scientific studies are going on not only on outer-space problems, medicine, industry, etc., but also in psychology and education. At present, there is much interest in how people think.
>
> The ________ Junior High School has been selected as one of the junior high schools near Philadelphia to take part in a scientific study of problem solving or thinking. The study is sponsored by the United States Office of Education and the University of Pennsylvania. It requires that students take a good many different kinds of tests, ranging from very easy tests to quite hard ones. To make the study valuable, all students must do their best on all tests.
>
> I know that all of you will do your best, both to insure a good study and to give yourselves practice in taking tests. When you apply for admission to college or for a job after you finish high school, you very likely will have to take a good many tests, many of which will be like the ones you are going to take during the study.
>
> The first four tests are being given this week. In about two weeks another set of tests will be given. Then about one-tenth of you will be selected partly by chance to take individual tests. All of the individual tests will be interesting and you will enjoy them, as well as the group tests. I wish all of you could take them, but they require too much time. But those of you who do take them can tell us about the interesting problems and puzzles and perhaps we can try some of them in class.
>
> Let me emphasize again that you must do your best. How you do on the tests will be confidential, and will appear only on your permanent records. These tests do not affect your grades, but they will be useful if you wish advice about what subjects to take in high school or what kind of occupations to prepare for.

* * *

DIFFERENTIAL APTITUDE TESTS, verbal, abstract, and mechanical reasoning subtests.

> These tests are accompanied by thorough directions in the manual, "Directions for Administration." We have indicated in red pencil particularly important parts of the directions. The tests should be given in the order indicated, and should be given on the same day. If they are given in one 90-minute period, brief rest periods should be allowed between the tests. The tests will be difficult for many seventh- and eighth-graders, and the idea of doing one's best should be stressed.

DAVIS-EELLS GAMES, exclusive of money problems (items 21-28)

This test is rather difficult to administer, particularly since we are using answer sheets instead of the spaces in the booklets. Follow the manual, "Directions for Administering and Scoring" exactly. The problems and options are to be read only once. The test will be easy for seventh- and eighth-graders, and can be given in one period of about 60 minutes. Be sure that pupils properly use answer sheets and that they do not make marks in the booklets.

STANDARD PROGRESSIVE MATRICES

This test is easy to administer. Just follow the photo-copy of "Instructions for Using Progressive Matrices." Be sure that pupils understand what they are to do, make proper use of the answer sheets, and do not put marks in the booklets. Allow 50 minutes, unless about 90 per cent of the pupils finish before that time. The only information needed on the answer sheet is name and school.

THOUGHT PROBLEMS

This test is divided in two parts, and the parts should be given on separate days. Sixty minutes should be allowed for each part, unless 90 per cent of the pupils finish before that time. The problems range from easy to quite difficult. The ideas of trying every problem and of showing figures or diagrams, where possible, should be stressed. In other words, stress the directions on the covers of the tests.

* * *

On the large data sheets which were earlier sent to principals for names and ranks of pupils in sections, please list the following:

Narrow	columns	1 and 2	Language IQ
"	"	4 and 5	Non-Language IQ
"	"	6 and 7	Total IQ
"	"	9 and 10	Reading achievement(percentile or grade norm)
"	"	12 and 13	Arithmetic achievement " " " "

On the slanted lines at the top, please indicate the names of the tests used and the dates given.

School of Education
University of Pennsylvania
Philadelphia 4, Pennsylvania
November 7, 1958

TO: Teachers who will administer tests for study of problem-solving ability.

FROM: Ethel Maw, Barbara Stanier, M. W. Tate

We have scored most of the first battery of tests. Thank you very much for the careful administration reflected in the answer sheets.

Our second battery consists of the tests shown on the following page. Some of these tests will be unfamiliar to you and your pupils, but we believe that all may bring out important skills associated with problem-solving ability, and we believe most will be interesting and easy to administer.

The tests and directions are largely self-explanatory. We have included time limits, nature of tests, and special comments on the following page. The general directions for giving the tests are:

1. Follow the order of tests worked out by your principal. (Although the tests may be given in any order in a school, with the exceptions noted on the following page, all sections must follow agreed-upon order at the scheduled time. This is because pupils may discuss tests with friends in other sections.)

2. Encourage pupils to do their best. (It might be a good idea to repeat parts of the general statements to pupils in our memorandum of October 17.)

3. Read the directions printed on each test aloud while your pupils read them silently. Repeat the directions if necessary, but do not supplement or modify.

4. See that pupils do not turn pages or begin tests ahead of time.

5. Make sure that pupils do not help each other.

6. Observe time limits as accurately as possible with second hand of an ordinary watch.

7. When you collect the papers be sure that pages are not separated. If staples have come loose, please "dog-ear" corners. Keep the tests together by sections.

Test[a]	Time Limit (min.)		Nature of Test	Comments for Administrator
	Per Test	Total		
1. Words[b]		3	Word fluency	Two copies for each pupil are provided; one for beginning and one for end of testing program.
2. First letter[b]	3	6	Word fluency	
3. First & last letters[b]	4	8	Word fluency	
4. Similar words[b]	4	8	Word fluency	
5. Memory		15[c]	Memory, interest	Give this test early in the first testing period.
6. Reasons	6	12	Idea fluency	
7. Ideas	5	10	Idea fluency	
8. Seeing problems	4	8	Sensitivity to problems	
9. Missing facts		35[c]	Comprehension of problems	
10. Planning		20[c]	Seeing needed facts	
11. Groups of things	3	6	Idea fluency	
12. Consequences	8	16	Originality, foresight	
13. Uses	4	8	Fluency, originality	Do not give on same day as"Seeing Problems"
14. Mutilated pictures		10	Discrimination of detail	
15. Hidden pictures		12	Flexibility of perception	
16. Incomplete pictures		5	Seeing wholes from parts	
17. Incomplete words		5	Seeing wholes from parts	Give tests 16 and 17 consecutively.
18. Concealed figures		12	Flexibility of perception	
19. Estimation		30[c]	Judgment	
20. Preferences		25[c]	Attitudes toward thinking activities.	

---

[a] Depending on the results of the scheduled tests, these additional tests may be given: practical judgment, critical thinking, location of error, generalization, difficulty reduction.

[b] Tests 1-4 are word association tests under condition which become progressively more restrictive. We believe that differences between good and poor problem solvers will become greater as conditions become more restrictive.

[c] If about 90 per cent of the pupils finish before time is up, stop the test